Essays on Econometrics and Policy Evaluation

by

Jaume Vives-i-Bastida

Submitted to the Department of Economics and the Statistics and Data Science Center
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS AND STATISTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

Authored by: Jaume Vives-i-Bastida
Department of Economics and the Statistics and Data Science Center
May 15, 2025

Certified by: Alberto Abadie
Professor of Economics, Thesis Supervisor

Certified by: Anna Mikusheva
Edward A. Abdun-Nur (1924) Professor of Economics, Thesis Supervisor

Accepted by: Isaiah Smith Andrews
Charles E. and Susan T. Harris Professor of Economics
Graduate Officer, Department of Economics

Essays on Econometrics and Policy Evaluation

by

Jaume Vives-i-Bastida

Submitted to the Department of Economics and the Statistics and Data Science Center on May 15, 2025 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS AND STATISTICS

ABSTRACT

This thesis consists of four chapters that study the statistical properties of synthetic control methods and their application to public policy evaluation and the digital economy.

The first chapter, co-written with Ahmet Gulek, proposes a Synthetic Instrumental Variables (SIV) estimator for panel data that combines the strengths of instrumental variables and synthetic controls to address unmeasured confounding. We derive conditions under which SIV is consistent and asymptotically normal, even when the standard IV estimator is not. Motivated by the finite sample properties of our estimator, we introduce an ensemble estimator that simultaneously addresses multiple sources of bias and provide a permutation-based inference procedure. We demonstrate the effectiveness of our methods through a calibrated simulation exercise, two shift-share empirical applications, and an application in digital economics that includes both observational data and data from a randomized control trial. In our primary empirical application, we examine the impact of the Syrian refugee crisis on Turkish labor markets. Here, the SIV estimator reveals significant effects that the standard IV does not capture. Similarly, in our digital economics application, the SIV estimator successfully recovers the experimental estimates, whereas the standard IV does not.

The second chapter, co-written with Ignacio Martinez, proposes a Bayesian alternative to the synthetic control method and explores the frequentist properties of the method in the context of linear factor models. In this chapter, we characterize the conditions on the factor model primitives (the factor loadings) for which the statistical risk minimizers are synthetic controls (in the simplex). Then, we propose a Bayesian alternative to the synthetic control method that preserves the main features of the standard method and provides a new way of doing valid inference. We explore a Bernstein-von Mises style result to link our Bayesian inference to the frequentist inference. For linear factor model frameworks we show that a maximum likelihood estimator (MLE) of the synthetic control weights can consistently estimate the predictive function of the potential outcomes for the treated unit and that our

Bayes estimator is asymptotically close to the MLE in the total variation sense. Through simulations, we show that there is convergence between the Bayesian and frequentist approach even in sparse settings. Finally, we apply the method to re-visit the study of the economic costs of the German re-unification and the Catalan secession movement. The Bayesian synthetic control method is available in the bsynth R-package.

The third chapter, recognizes that synthetic control methods often rely on matching pre-treatment characteristics (called predictors) of the treated unit, and that the choice of predictors and how they are weighted plays a key role in the performance and interpretability of synthetic control estimators. This chapter proposes the use of a sparse synthetic control procedure that penalizes the number of predictors used in generating the counterfactual to select the most important predictors. I derive, in a linear factor model framework, a new model selection consistency result and show that the penalized procedure has a faster mean squared error convergence rate. Through a simulation study, I then show that the sparse synthetic control achieves lower bias and has better post-treatment performance than the unpenalized synthetic control. Finally, I apply the method to revisit the study of the passage of Proposition 99 in California in an augmented setting with a large number of predictors available.

The fourth chapter, co-written with Alberto Abadie, proposes a set of simple principles to guide empirical practice in synthetic control studies. The proposed principles follow from formal properties of synthetic control estimators, and pertain to the nature, implications, and prevention of over-fitting biases within a synthetic control framework, to the interpretability of the results, and to the availability of validation exercises. We discuss and visually demonstrate the relevance of the proposed principles under a variety of data configurations.

JEL: C23, C26, C11, C52.

Thesis supervisor: Alberto Abadie
Title: Professor of Economics

Thesis supervisor: Anna Mikusheva
Title: Edward A. Abdun-Nur (1924) Professor of Economics

Acknowledgments

I am profoundly grateful to my advisors Alberto Abadie, Anna Mikusheva and Tobias Salz for their invaluable support during the PhD.

Alberto's guidance since the start of the PhD has shaped me as a researcher and as a person. Our frequent meetings, collaborations, and discussions have been the backbone of my PhD years, and I will remember them fondly. I have learned from Alberto the importance of focusing on empirically relevant problems as an econometrician and the craft of writing econometric papers. Beyond the academic, I will always be in debt to Alberto for his support during the PhD at a personal level. His warm approach and care have carried me through the hard and good times of the PhD. Thank you Alberto!

I will always remember Anna's kindness and careful eye as the right way to overcome research roadblocks. Thank you Anna for all your help working through the projects during the PhD, for all the meetings from which I always left happy and with a higher sense of clarity, and for being there for me in the stressful times.

While my thesis ended up gravitating more towards Econometrics than Industrial Organization, I have benefit tremendously during my PhD from my discussions with Tobias. From platform economics and recommendation systems, to applied econometrics and structural modeling, thank you Tobias for your advise and enthusiasm during the PhD and for the candor of our conversations about research and life.

Beyond my advisors, I am also thankful to everyone in the Department of Economics and the Institute for Data, Systems, and Society at MIT for their availability and support, and to numerous researchers that have provided feedback on the projects in this thesis. I am especially grateful to Whitney Newey, Isaiah Andrews, Victor Chernozhukov, Max Kasy, Rahul Singh, Liyang Sun, and Jared Greathouse among many others, for helpful comments and conversations during the PhD.

One of the greatest parts of the PhD has been collaborating with my co-authors; Ahmet Gulek, Alejandro Sabal, Ignacio Martinez, Sylvia Klosin, Judit Vall, Nuria Mas, Jon Gruber, Chinmay Lohani and Alberto Abadie. Thank you for your patience and collegiality while working with me, I look forward to future projects.

Above all, this thesis, and the PhD journey it represents, would not have been possible without the incredible people who have supported and accompanied me over the years. My deepest thanks go to Pier, whose friendship has been the highlight of my time in Boston, and to Bea and Coti, for the unforgettable years we shared in the Highlands. I am also deeply grateful to my office mates; Adam, Rafa, Sylvia, Marc, Bumsoo, Nancy, Anne, and Nagisa, for making the long PhD days not only bearable but often enjoyable. Beyond MIT, I've been fortunate to share this journey with amazing friends. Thank you Alejandro, Pep, Xita, Santi, Marc C., Marc T., Albert, and Chris, among many others. Finally, I owe my greatest gratitude to my family; my parents, Xavier and Aurora, my grandparents, Leandre and Tona, and my younger brother, Martí; I dedicate this dissertation to all of them.

Contents

List of Figures

List of Tables

Chapter 1

Synthetic IV Estimation in Panels

written jointly with Ahmet Gulek

1.1 Introduction

In this paper, we propose a Synthetic Instrumental Variables (SIV) estimator that combines instrumental variables and synthetic controls to account for bias due to unobserved confounding. We are interested in panel data settings in which an intervention affects a set of units over time, but we are worried about endogeneity concerns such as the intervention affecting units selectively or differential trends amongst units that received different doses of the treatment. In this context, researchers may turn to differences-in-differences (DiD) designs (Card and Krueger, 2000) or synthetic control (Abadie and Gardeazabal, 2003) designs (SC) in which control units are used to evaluate the counterfactual in absence of the intervention. While these approaches may address part of the endogeneity problems, often valid control units may not exist, as all units may be treated, or control units and treated units may not follow similar paths, violating the parallel trends assumption. Faced with this challenge, we may consider an instrumental variable (IV) approach in combination with the DiD design (for example using a shift-share instrument, e.g. Jaeger et al. (2018)). In practice, however, the endogeneity concerns may persist as the instrument may be correlated with unobserved confounders in the outcome of interest. The SIV estimator provides a solution to this problem.

To understand the relevance of our setting, consider the use of shift-share instrumental variables (SSIV) to identify causal effects of a treatment or policy by comparing groups or regions more and less exposed to the treatment. Influential examples include studies of the effects of immigration (Card, 2001) and trade (Autor et al., 2013) on labor markets. In this paper, our main empirical application concerns the study of the effect of immigration on Turkish labor markets using the Syrian civil war as an exogenous shock. An intuitive empirical strategy to address the endogeneity of immigrants' location choice, is to use a shift-share design where the distance to the border is the "share" and the aggregate inflow of immigrants is the "shift." Identification in this context relies on regions close to and away from the border to follow parallel trends absent migration flows. The problem with such a design is that regions close and away from the border may be on different economic trajectories before the Syrian civil war starts. These differential trends may bias our estimates of the effect of the refugees on local labor outcomes. Our proposed method, the SIV, creates a synthetic control unit for each region in the pre-intervention period and then debiases the outcomes of interest to account for the differential trends and correct the bias in the two-stage least square estimator (TSLS).

We motivate our method theoretically by deriving consistency and asymptotic normality results in triangular panel designs with unmeasured confounding. We assume that the unobserved error term has two components: an idiosyncratic component that is orthogonal to the instrument and an unobserved heterogeneity component that follows a factor structure. If we could control for the unobserved factor structure the TSLS would be consistent, but we cannot do so directly. Our solution, the SIV, proposes synthetic controls as a way to proxy for the unobserved confounding through interpolation. Under signal-to-noise restrictions and weak primitive assumptions we show that the synthetic IV is consistent and asymptotically normal when the number of units and time periods is large. Through finite sample bounds we highlight that the proposed estimator might be especially sensitive to the noise level and the weakness of the instrument. To guard against small sample biases of the estimator, we propose empirical checks researchers might want to implement in practice, as well as a "doubly robust" ensemble estimator that combines the synthetic IV with a projected synthetic IV that partials out the noise. We also provide an alternative permutation based inference procedure that is exactly valid in small samples.

We show the applicability of our method in a calibrated simulation exercise, by studying the Syrian refugee crisis example, by re-visiting the effect of Chinese imports on US manufacturing employment (Autor et al. (2013)) and by studying the effect of producer rankings on sales in digital platforms. The simulation study shows that the synthetic IV and ensemble estimators outperform the TSLS (with two-way fixed effects) in a variety of settings. Furthermore, the SIV exhibits close to zero bias in settings with moderate and small levels of noise and unmeasured confounding, and the ensemble estimator is shown to be robust in settings with higher noise levels. In a study of the coverage of the synthetic IV estimator confidence intervals we find that it is good in cases in which the estimator exhibits small bias. Following the theoretical properties and the observed behavior under simulations we recommend that researchers implement four checks (in the spirit of the best practices detailed in Abadie and Vives-i-Bastida (2025)) when using the estimator: (1) ensuring that the instrument is not weak after the debiasing, (2) making sure that the estimator achieves good fit in the pre-treatment period, (3) implementing a back test to ensure the good fit is not due to over-fitting to the idiosyncratic noise and (4) ensuring the synthetic controls weights are dense, with no one unit receiving all the weight.

In our study of the effects of Syrian migrants on Turkish labor markets we find that regions close to and away from the border follow different trajectories in the pre-period, potentially biasing the estimates from a shift-share design. However, the SIV corrects for this problem and the debiased estimates do not exhibit pre-trends. Moreover, using SIV leads to different conclusions, relative to the standard shift-share IV, about the effect of immigration in the Turkish context. While the shift-share IV estimator cannot reject that there is no effect of immigration on natives' salaried employment, the synthetic IV estimator finds a statistically significant negative effect. For example, using SSIV we find that a 1 percentage point (pp) increase in refugee/native ratio is associated with a 0.01 pp *increase* in native salaried employment for low-skilled men, whereas using SIV we find that it causes a, statistically significant, 0.16 pp *decrease*. This implies that for every 100 immigrants that arrived to Turkey, 16 low-skilled natives lost salaried jobs. These economically and statistically significant differences between the SSIV and SIV estimates highlight the role of unobserved confounders in the long-standing debate about the labor market effects of immigrants (Borjas, 2017; Peri and Yasenov, 2019).

In our re-analysis of the effect of Chinese imports on US manufacturing, we follow the identification strategy of Autor et al. (2013). We compare regions that were more exposed to Chinese imports based on their pre-existing industrial composition with regions that were less exposed. We first show that the "shares" are correlated with regional growth rates in the pre-period. Regions that were more exposed to the China shock starting from 1990 grew less in 1970s and 1980s. In fact, the difference in growth rates in the 1970s and 1980s is almost identical to the difference in growth rates after the China shock in 1990s and 2000s. The SIV estimator corrects for this pre-trend, and finds smaller effects in the 1990s than the original SSIV, but similar effects in the 2000s. This evidence implies that regional trends between 1970–1990 account for about half of the effect the standard SSIV captures in 1990s. This is intuitive as the economic trajectories between 1970–1990 are more likely to continue in 1990s than 2000s. Our findings contribute to the growing literature estimating the China shock effect under different modelling assumptions (Goldsmith-Pinkham et al., 2020).

Finally, we apply the SIV to study an important question in digital economics: how much do rankings affect producer outcomes? In the context of a large food delivery platform we use preferential contracts given to producers that mechanically increase their ranks in the consumer search wall as an IV for their rank. In this context, we have at our disposal an A/B test in which rank was randomized which allows us to benchmark our observational estimates on the effect rank. As expected, we find that the standard IV (with two-way fixed effects) exhibits positive omitted variable bias (relative to the A/B test) as producers that receive the preferential contracts are in an upwards trend relative to others. The synthetic IV estimates however do not exhibit such bias and recover the A/B test estimates. This examples corroborates the usefulness of our proposed estimator in dealing with unmeasured confounding in IV-DiD settings and provides a new strategy to measure the causal effect of rank in digital platforms using instrumental variables. Given the challenges associated with IVs in this context (as discussed by Rutz et al. (2012)) we see this as a contribution to the literature in digital economics.

This paper contributes to several strands of the literature. First, it complements the growing body of work on addressing unobserved confounding and 'pre-trends' in panel data settings by providing a new method for the IV DiD case. Research in this area is built upon synthetic control based methods (Abadie et al., 2010, 2015; Ben-Michael et al., 2021; Imbens and Viviano, 2023), more general weighting methods such as the synthetic differences in differences (Arkhangelsky et al., 2021), as well as balancing methods (Hainmueller, 2012), matrix completion methods (Agarwal et al., 2021b; Athey et al., 2021) and factor model methods (Anatolyev and Mikusheva, 2022; Bai, 2009). Similarly, our paper complements related work on addressing and evaluating pre-trends in event-study designs, including Freyaldenhoven et al. (2019), Borusyak et al. (2023), Roth (2022) and Ham and Miratrix (2022) among others. A more closely related paper is Arkhangelsky and Korovkin (2023) which provide a novel weighting algorithm to address unobserved confounding in settings in which the exogenous variation comes from aggregate time series shocks. The authors propose a robust estimator that corrects the TSLS bias when the instrument has a product form and there are unobserved aggregate shocks that may affect different units differently. We see our method as complementary to Arkhangelsky and Korovkin (2023), and note that we consider a different setting in which the instrument need not have a product structure and the exogenous variation may come from the time or unit components.

Second, this paper is related to a growing literature studying and relaxing the identification assumptions embedded in shift-share designs. Goldsmith-Pinkham et al. (2020) show that the identification assumptions in SSIV designs are often based on the exogeneity of shares. Borusyak et al. (2022) relax this assumption and provide a framework in which identification can also come from the exogeneity of shifts, allowing shares to be endogenous. Adao et al. (2019) highlight an inference problem that arises from cross-regional correlation in the regression residuals due to similarity of sectoral shares in the US. de Chaisemartin and Lei (2023) investigate a weighting problem in SSIV with heterogeneous treatment effects and propose a robust correlated-random-coefficient panel IV estimator. In the immigration context, Jaeger et al. (2018) show that past-settlement instruments in practice conflate both short-term and long-term adjustments to immigration shocks, which invalidates the exogeneity assumption. Our method complements this literature by providing a new tool applied researchers can rely on to address unobserved confounding in SSIV designs.

Lastly, our main empirical example is related to a large literature studying the effects of immigration using refugee shocks (Card, 1990; Hunt, 1992; Friedberg, 2001; Angrist and Kugler, 2003; Lebow, 2022). More specifically, our focus on the effects of Syrian refugees on Turkish natives and the presence of unobserved confounders in Turkey follows Gulek (2025). Whereas he focuses on the effects on the formal and informal labor markets, we focus on the overall impact on salaried employment and consider heterogeneity across men and women.

The paper proceeds as follows. Section 1.2 describes the setting and an empirical example. Section 1.3 presents the synthetic IV estimator and two additional estimators. Section 1.4 discusses the theoretical results. Section 1.5 regards extensions of the estimator and inference procedures. Section 1.6 concerns the simulation study, and section 1.7 details our three empirical applications.

1.2 General setting and empirical motivation

We are interested in a panel data setting in which some units of interest are exposed to a (potentially continuous) treatment and there are endogeneity concerns. The researcher may be worried about using a differences-in-differences design as the parallel trends assumption might not hold, but has access to an instrument that *partially* addresses the endogeneity concerns. More precisely, we consider J units indexed by $i = 1, \ldots, J$ that are observed for T periods of time with outcomes of interest Y_{it} and potential outcomes denoted by $Y_{it}(R_{it})$ for a random variable $R_{it} \in \mathbb{R}$. Throughout the paper we assume that the potential outcomes are generated as described by the following assumption.

Assumption 1 (Design)**.** *Outcomes follow*

$$Y_{it}(R_{it}) = \theta R_{it} + U_{it} + \epsilon_{it}$$

where $Y_{it}, R_{it},$ and Z_{it} are observed and $U_{it} = \mu_i' F_t$, for a $k \times 1$ vector of factor loadings μ_i and a vector of common factors F_t, and ϵ_{it} are unobserved. The treatment R_{it} follows

$$R_{it} = \gamma Z_{it} + A_{it} + \eta_{it},$$

where Z_{it} is an instrument satisfying that $Z_{it} = 0$ for $t \leq T_0$, A_{it} is an unobserved heterogeneity term and η_{it} is an idiosyncratic shock potentially correlated with the error term ϵ_{it}.

The main feature of the panel triangular design we consider in Assumption 1 is that the instrument Z_{it} is not active for $t \leq T_0$ and the unobserved components are additively separable into unobserved heterogeneity components (U_{it} and A_{it}) and idiosyncratic error components (ϵ_{it} and η_{it}). The design reflects settings in which an observed intervention starts at T_0 and is used as an instrument or to construct an instrument Z_{it}. For example, shift-share designs satisfy this setting as $Z_{it} = Z_i' H_t$ for shares Z_i and shifts H_t with $H_t = 0$ for $t \leq T_0$. This will be the case for our application to the China shock study. Panel instrumental variable designs in which the instrument becomes active after T_0 (i.e. $Z_{it} = 0$ for $t \leq T_0$) also satisfy this setting, and this will be the case of our digital economics application. A special instance of our design is $R_{it} = 0$ for $t \leq T_0$ (i.e. $A_{it} = \eta_{it} = 0$ for $t \leq T_0$), which is satisfied in our main empirical application to the Syrian refugee crisis, as well as for common frameworks considered in the literature of IV-DID.

The parameter of interest is the expected marginal effect of the treatment R_{it} on the outcome Y_{it},

$$\theta = \mathbb{E}\left[\frac{\partial Y_{it}(R_{it})}{\partial R_{it}}\right] = \frac{\partial Y_{it}(R_{it})}{\partial R_{it}}.$$

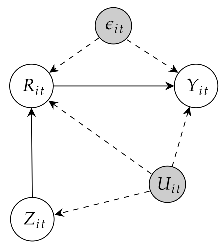
To understand the potential problems arising in the estimation of θ; it is useful to consider the possible identifying assumptions researchers may posit. To this end Figure 1.1 describes different assumptions on our design encoded in directed graphs. In this paper, we relax the standard IV independence assumption by allowing the instrument to be correlated with the unobserved term U_{it} as described in panel (a) of Figure 1.1 and Assumption 2.

Assumption 2 (Partial instrument exogeneity)**.** *The following independence conditions hold* $\epsilon_{it}, \eta_{it}, A_{it} \perp Z_{it}$.

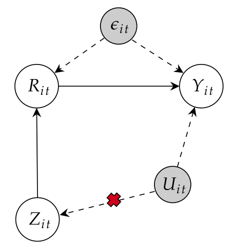
Tu put in context Assumption 2 we consider alternative assumptions researchers may posit. Researchers may consider the independence assumption $R_{it} \perp \epsilon_{it}, U_{it}$ (panel (c) in Figure 1.1) which in our design is satisfied when $A_{it}, Z_{it}, \eta_{it} \perp \epsilon_{it}, U_{it}$ pairwise. This assumption is not implied by Assumption 2, but if it holds, then the OLS estimator of θ is unbiased and the researcher could recover θ by regressing Y on R. In many empirical settings, however, R_{it} is likely correlated with the unobserved components. For example, in immigration settings refugees might take into account local labor market conditions and trends when choosing where to re-locate or alternatively may relocate based on geographical distance. Researchers may therefore rely on instrument Z_{it} to address this concern. Common instruments in the immigration literature to address location choice endogeneity include past-settlement indicators or travel distance (Card, 2001; Angrist and Kugler, 2003).

A valid instrument requires that (1) the only channel affecting the outcome Y is through the treatment R (the exclusion restriction) and (2) that the instrument is as good as randomly assigned. In our design, this would require that $\epsilon_{it}, \eta_{it}, U_{it}, A_{it} \perp Z_{it}$, which is also not implied by Assumption 2, as reflected in panel (b) in Figure 1.1. We are interested in cases in which the instrument Z_{it} is not perfectly valid due to failing to be as good as randomly assigned. This may be the case in some relevant empirical settings. For instance, in immigration examples regions that received immigrants in the past or were closer to the immigrants' origin may follow different trends than other regions. Given that models of migrant location decisions (Llull (2017), Bartel (1989)) often involve agents taking expectations over both settling costs and the economic returns of settling in a particular region; it is likely that a single instrument Z_{it} may not simultaneously address both sets of endogenous variables.

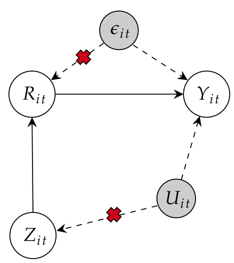
Figure 1.1: Triangular designs



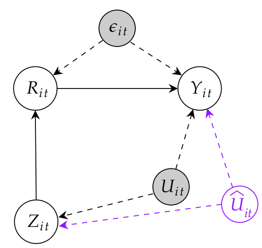
(a) General design



(b) IV design



(c) OLS design



(d) Proxy design

Notes: Directed acyclic graphs representing the independence assumptions implicit for different designs. Variables shaded grey are unobserved.

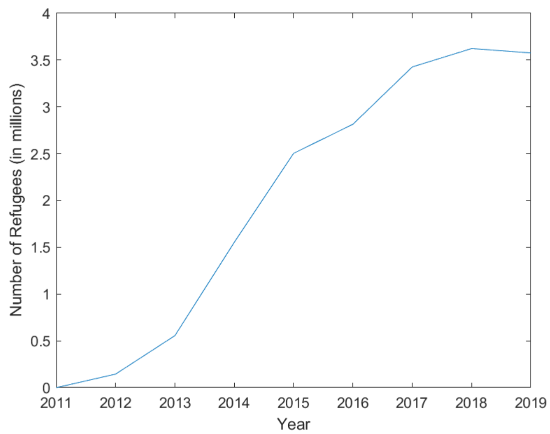
However, we may believe that the instrument Z_{it} is valid for one set of variables (settling costs for example) and that the design would be valid if we could control for economic trends. In other words, the researcher may believe there exists an omitted variable that is correlated with the instrument and the outcome of interest. This motivates us to relax the instrument independence conditions as posited in Assumption 2 to distinguish between the unobserved component ϵ_{it} unrelated to the instrument from the unobserved heterogeneity component U_{it}.

Assumptions 1 and 2 imply that the instrument Z_{it} correctly addresses the endogeneity problem due to the unobserved component ϵ_{it}, but not the omitted variable bias due to U_{it}. If we could observe U_{it}, we would control for it and the IV design would be valid. However, U_{it} is unmeasured and, therefore, the statistical problem we consider is that of finding a valid proxy control $\hat{U}_{it}$ for U_{it} (as depicted in panel (d) of Figure 1.1). Different approaches have been considered in the literature to tackle this problem. If additional variables are available, strategies have been proposed to combine the observed variables and the instrument to proxy for U_{it} directly (see Miao et al. (2018) and Deaner (2021)). If there exists a donor pool of units never exposed to the treatment R, researchers may opt for a different design and use synthetic controls to partial out U_{it} (Cengiz and Tekgüç, 2022). However, in many empirical settings additional variables or additional control units are not available, and in these cases researchers often rely on parametric assumptions to control for U_{it}. Common examples used in the literature include two-way fixed effects, $U_{it} = \alpha_i + \delta_t$, linear time trends (Wolfers, 2006), $U_{it} = \alpha_i \times t$, or grouped region time fixed effects (Stephens Jr and Yang, 2014; Bonhomme and Manresa, 2015), $U_{it} = \alpha_g \delta_t$ for region groups g. The choice of parametric form is often driven by domain knowledge, or, when possible, by showing that the instrument Z_{it} is not correlated with the outcome in the pre-treatment period (see Danieli et al. (2024) for a review of IV falsification tests).

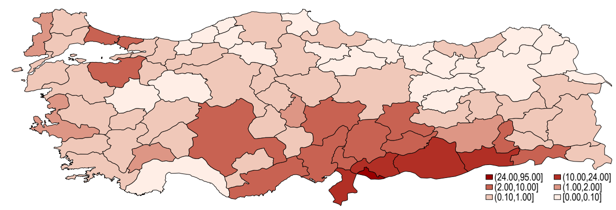
In this paper, we propose a strategy to directly use the pre-treatment period ($t \leq T_0$) to flexibly control for U_{it} without estimating a specific functional form. Our proposed method, the *synthetic* IV, uses the pre-treatment period to estimate synthetic control weights that interpolate across units to partial out U_{it} in the post-treatment period ($t > T_0$). The idea of using the pre-treatment period as a way to address an omitted variable bias in panel IV estimates is not completely novel, and this is often done in IV-DID settings by instrumenting the difference of outcomes before and after treatment (sometimes called the differences in Wald estimator). However, the validity of this approach still relies on strong parametric assumptions on U_{it}, for example Danieli et al. (2024) show for that for a binary setting under the parallel trends assumption (which implicitly imposes that $U_{it} = \alpha_i + \delta_t$), the differences in Wald estimator is an unbiased estimator for θ. The method we propose works for continuous treatments and instruments, and is valid under more general functional forms; linear factor models $U_{it} = \mu_i' F_t$. Given that many economic trend variables can be described through an interactive factor structure we believe that our method can be quite useful to researchers to flexibly control for unmeasured confounding and prevent the specification search over different fixed effect combinations. Furthermore, linear factor model assumptions are common in the literature for dealing with unmeasured confounding in panel settings. A related paper that also relies on a factor model structure is Arkhangelsky and Korovkin (2023) in which an aggregation scheme based on Arkhangelsky et al. (2021) is used to control for unmeasured aggregate confounders between a time series instrument Z_t and a panel outcome Y_{it}. We see our paper as complementing the novel work of Arkhangelsky and Korovkin (2023) for cases in which we have a panel instrument Z_{it} and our identifying assumptions may come from unit level variation in the instrument or from time series variation, which include the commonly used shift-share designs.

Example 1 (the Syrian refugee crisis) To further motivate why the setting described under Assumptions 1-2 and in Figure 1.1 is relevant to applied work, consider our main

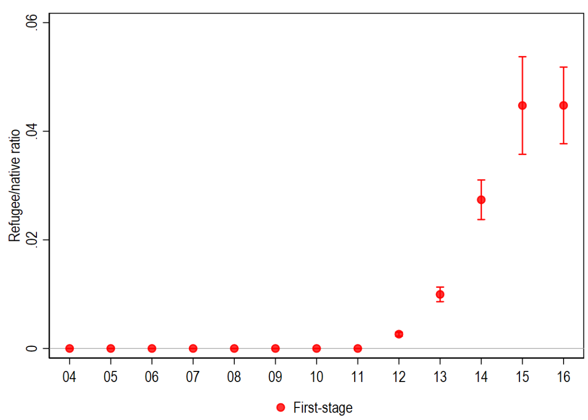
Figure 1.2: The Syrian refugee shock



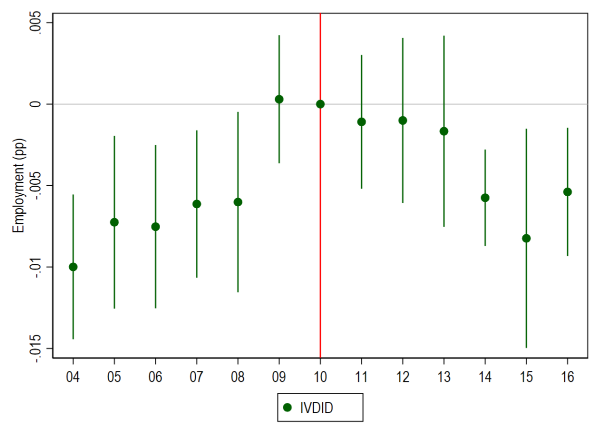
(a) Number of refugees



(b) Refugee shares



(c) First stage



(d) Reduced form

Notes: In event-study designs the 95% confidence intervals are plotted. The F-stat of the main first-stage regression is 154. In Panels (c) and (d) the x axis shows the years 2004-2016 in 2 digit notation.

empirical example: the effect of the Syrian refugee crisis on Turkey's local labor markets. The Syrian civil war started in March 2011 and by 2017, 6 million Syrians had sought shelter outside of Syria with 3.5 million locating in Turkey.[1] Figure 1.2 panel (a) shows the growth in the number of Syrian refugees in Turkey over time and panel (b) shows the geographic dispersion of the refugees. Given the structure of the Syrian refugee shock a natural approach to estimating the impact of refugees on local labor outcomes is that of a shift-share instrumental variable design that exploits the exogenous time shock of the civil war and the differential impact across units.

To relate the Syrian example to our setting let R_{it} denote the refugee/native ratio at

[1] Turkey hosts the largest number of refugees in the world (UNHCR, 2021).

province-year level and consider a travel distance shift-share instrument, as is common in the mass-immigration literature (Angrist and Kugler, 2003; Aksu et al., 2022).

$$Z_{it} = \underbrace{\bar{H}_t}_{\text{shift}} \times \underbrace{Z_i}_{\text{share}},$$
$$Z_i = \sum_{s=1}^{13} \lambda_s \frac{1}{d_{i,s}}$$

where $\bar{H}_t$ is the number of refugees in Turkey in year t, $d_{i,s}$ is the travel distance between Turkish region i and Syrian governorate s, λ_s is the weight given to Syrian governorate s which we set it be proportional to the population share of s.[2] In panel (c) of Figure 1.2 we plot the first stage coefficients interacted with time dummies from the TWFE specification that is commonly estimated in the literature

$$R_{it} = \sum_{k \neq 2010} \gamma_k (\mathbb{1}\{t = k\} \times Z_i) + \alpha_i + \delta_t + \eta_{it}.$$

The first stage regression tests whether the instrument predicts refugees' location choice every year. As expected, distance is a strong predictor of the refugee treatment intensity. The F-stat of the shift-share first-stage (where we regress R_{it} on Z_{it} while controlling for region and time f.e.) is 154. The problem arises when one considers the reduced form of local wage-employment (salaried employment) of the natives that did not finish high-school (low-skill)[3] on the instrument

$$Y_{it} = \sum_{k \neq 2010} \theta_k (\mathbb{1}\{t = k\} \times Z_i) + \alpha_i + \delta_t + \epsilon_{it}, \tag{1.1}$$

which is displayed in panel (d) of Figure 1.2. Between 2004–2010 (before the refugee crisis began), the provinces closer to the border observed employment gains compared to other regions. Being one standard deviation closer to the border predicts a wage-employment growth of 1 pp between 2004 and 2009. Given that the regions that are predicted by the instrument to receive immigrants were following different trends *before* the shock, it is likely that the IV-DID design does not satisfy the parallel trends assumption implicit in the TWFE specification. This suggests that there exists an unmeasured confounder U_{it} in the Syrian crisis empirical setting. The appearance of pre-trends in similar designs is a common problem

[2]The idea is that all else equal, more Syrians would be expected to come from the more populous regions.

[3]This is the key outcome of interest because Syrian refugees were substantially less educated compared to the Turkish population, and hence constitute largely a low-skill immigration shock. We provide more details about the setting in the Appendix.

in practice (Wolfers, 2006; Stephens Jr and Yang, 2014; Gulek, 2025) and has been discussed extensively in the literature (Roth, 2022; Freyaldenhoven et al., 2019). While our main design described by Assumption 1 considers a common, time-invariant, parameter θ, in the Appendix (section A.1.7) we extend it to a time-varying event study design like the one considered in this example.

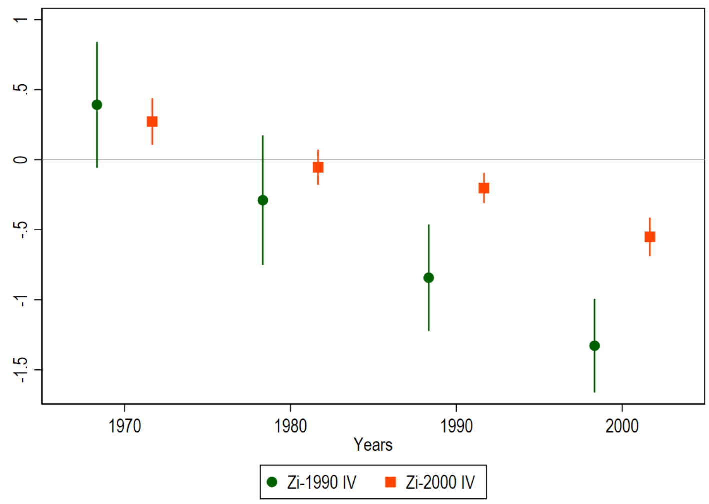
Example 2 (the China shock) Another shift-share research design that highlights the problem of interest is used by Autor et al. (2013) to evaluate the effect of Chinese exports on manufacturing employment in the US. The authors are interested in the effect of US import's exposure to China on the growth in percent manufacturing employment (Y_{it}) for a US commuting zone i during a decade t. Given that import exposure is endogenous, Autor et al. (2013) instrument it by the increase in Chinese imports by high-income countries (Z_{it}). It is important to note that the instrument has a shift-share structure. Therefore, we can interact the exposure components in the instrument with time indicators to see whether they predict changes in the outcome *before* the shock occurs as we did for the Syrian example. In particular, we estimate the following reduced form regression

$$Y_{it} = \sum_k (\bar{Z}_{i,h} \times \mathbb{1}\{t = k\})\beta_{k,h} + \delta_t + \epsilon_{it} \tag{1.2}$$

where δ_t is a time fixed effect, ϵ_{it} is an error term, and $\beta_{k,h}$ with $h = 1990, 2000$ are the event-study estimates of interest for the two exposure measures $\bar{Z}_{i,h}$ considered by the authors in constructing the instrument Z_{it}. In section 1.7.2 we explain in detail how the shift-share instrument is constructed and replicated the main tables of Autor et al. (2013). Note, that we do not include region fixed effects following the original paper, but in principle we could at the cost of having to normalize one of the pre-period estimates to zero.

Figure 1.3 shows the event studies estimates for the reduced form regression (1.2). Each data point represents a coefficient estimate for a given decade, with the intervention (the China shock) occurring in the decade of 1990-2000. As it can be seen, for both exposure measures it appears that before the intervention regions more affected by Chinese imports where potentially on a downward trend in terms of manufacturing employment, raising concerns over the exogeneity of shares assumption implicit in the shift-share design. This example highlights that unmeasured confounding may be a common problem in shift-share designs. Autor et al. (2013) partially address these concerns by using additional covariates to control for the differential trends. In section 1.7.2 we revisit this example and show how our proposed method can correct for the unmeasured confounding without relying on additional data.

Figure 1.3: Reduced-form estimates for the China shock



Notes: Event-study estimates for (1.2) using the exposure components $\bar{Z}_{i,1990}$ and $\bar{Z}_{i,2000}$. The time periods consist of four decades 1970-1980, 1980-1990, 1990-2000 and 2000-2007, with the intervention (the China shock) occurring in the decade of 1990-2000.

In the following section we describe our proposed solution, the synthetic IV estimator, that flexibly controls for the unmeasured confounder U_{it}.

1.3 The synthetic estimator

The synthetic estimator consists of two steps. In the first step we find synthetic controls for each unit in a pre-period ($t < T_0$) and generate counterfactual estimates for Y_{it}, R_{it} and Z_{it} for a post period. In the second step, as in the standard IV estimator, we use these counterfactual estimates to compute the first stage and reduced form estimates. To describe the procedure, consider J units indexed by $j = 1, \dots, J$ observed for T periods of time. We are interested in an outcome of interest Y_{it} with potential outcomes $Y_{it}(R_{it})$ indexed by random variable R_{it}.

Step 1: for each $j \in \{1, \dots, J\}$ we find the synthetic control weights $\hat{w}_j^{SC}$ by solving the following program for the pre-treatment period $t \in \{1, \dots, T_0\}$

$$\hat{w}_j^{SC} \in \text{argmin}_{w \in \mathcal{W}} \| D_j^{T_0} - D_{-j}^{T_0 \prime} w \|^2, \tag{1.3}$$

for

$$\mathcal{W} = \{ w \in \mathbb{R}^J \mid \|w\|_1 \leq C \}, \tag{1.4}$$

where $C \in (0, \infty)$ is a regularization hyper-parameter, D^{T_0} is the $J \times p$ design matrix that includes pre-treatment outcomes Y_{jt} and treatments R_{it} for $t < T_0$ where $p = 2T_0$, with $D_j^{T_0}$ denoting the predictors for unit j and $D_{-j}^{T_0}$ the $(J-1) \times p$ matrix of predictors for the other units.[4] The intuition for matching outcomes and treatments is that the finite sample behavior of the estimator will depend on the pre-treatment fit of R_{it} and Y_{it}. In the case in which R_{it} is not present in the pre-period, as is the case of the Syrian refugee example (1.2), the design matrix includes only the pre-treatment outcomes, such that $D^{T_0} = Y^{T_0}$. The l_1 norm constraint on the weights ensures that there is some amount of regularization. This program is a relaxation of the standard synthetic control objective, sometimes called the constrained lasso (Doudchenko and Imbens, 2016). In our empirical applications, we compute the weights using the standard synthetic control restriction that the weights are in the simplex (i.e. $\mathcal{W} = \{w \mid w_j \geq 0, \quad \sum_j w_j = 1\}$). Our theoretical results will be valid for this case when $C = 1$ is chosen in solving program (1.3). We find in the simulations, and empirical exercises, that the additional regularization provided by the simplex restrictions offers good finite sample performance.

Once the synthetic control weights are computed, we define the following quantities for all t in $1, \ldots, T$

$$
\begin{aligned}
\hat{Y}_{it}^{SC} &= \sum_{j \neq i} \hat{w}_{ij}^{SC} Y_{jt}, \\
\hat{R}_{it}^{SC} &= \sum_{j \neq i} \hat{w}_{ij}^{SC} R_{jt}, \\
\hat{Z}_{it}^{SC} &= \sum_{j \neq i} \hat{w}_{ij}^{SC} Z_{jt},
\end{aligned}
$$

which we label the synthetic outcome, treatment level and instrument respectively. Then, we define the *debiased* values for $t > T_0$ as the difference between the observed values and the synthetic values

$$
\begin{aligned}
\tilde{Y}_{it} &= Y_{it} - \hat{Y}_{it}^{SC}, \\
\tilde{R}_{it} &= R_{it} - \hat{R}_{it}^{SC}, \\
\tilde{Z}_{it} &= Z_{it} - \hat{Z}_{it}^{SC}.
\end{aligned}
$$

Step 2: Given $\{\tilde{Y}_{it}, \tilde{R}_{it}, \tilde{Z}_{it}\}_{t=T_0+1}^{T}$, we estimate the first stage and reduced form by pooled

[4]Researchers may choose to weight the columns of the design matrix D^{T_0} according to different weights, as proposed by Abadie et al. (2010).

OLS regression

$$\tilde{\pi} \in \arg\min_{\pi} (\tilde{Y} - \tilde{Z}\pi)'(\tilde{Y} - \tilde{Z}\pi),$$
$$\tilde{\beta}_1 \in \arg\min_{\beta} (\tilde{R} - \tilde{Z}\beta)'(\tilde{R} - \tilde{Z}\beta).$$

where $\tilde{Y} = vec(\tilde{Y}^T)$, $\tilde{R} = vec(\tilde{R}^T)$ and $\tilde{Z} = vec(\tilde{Z}^T)$ are the $J(T - T_0) \times 1$ vectors of the debiased values. Then, the estimated average marginal effect is given by the standard IV estimate

$$\tilde{\theta}^{SIV} = \frac{\tilde{\pi}}{\tilde{\beta}_1},$$

which we denote the *synthetic* IV estimator. Given that our framework is a just-identified IV design, the IV estimator is equivalent to the two-stage least-squared estimator (TSLS) given by

$$\tilde{\theta}^{TSLS} = \left(\sum_{it} \tilde{Z}_{it}\tilde{R}_{it}\right)^{-1} \sum_{it} \tilde{Z}_{it}\tilde{Y}_{it}.$$

Our main asymptotic results are also valid for the estimator that uses the instrument Z_{it} instead of the de-biased instrument $\tilde{Z}_{it}$

$$\tilde{\theta}^{TSLS}_{YR} = \left(\sum_{it} Z_{it}\tilde{R}_{it}\right)^{-1} \sum_{it} Z_{it}\tilde{Y}_{it},$$

and the estimator that only debiases the instrument Z_{it}

$$\tilde{\theta}^{TSLS}_{Z} = \left(\sum_{it} \tilde{Z}_{it}R_{it}\right)^{-1} \sum_{it} \tilde{Z}_{it}Y_{it}.$$

In the theory and simulation sections we show that while these estimators are similar to the proposed *synthetic* IV estimator (SIV), they may have worse finite sample properties. To understand the differences between the standard IV and the SIV, we expand the *debiased* variable $\tilde{Y}_{it}$ in terms of $\tilde{R}_{it}$

$$\begin{aligned} \tilde{Y}_{it} &= Y_{it} - \hat{Y}^{SC}_{it} \\ &= \theta R_{it} + \mu_i' F_t + \epsilon_{it} - \sum_{j\neq i} \hat{w}^{SC}_{ij} Y_{jt} \\ &= \theta \tilde{R}_{it} + (\mu_i - \sum_{j\neq i} \hat{w}^{SC}_{ij}\mu_j)' F_t + \epsilon_{it} - \sum_{j\neq i} \hat{w}^{SC}_{ij}\epsilon_{jt}. \end{aligned} \tag{1.5}$$

It follows that the *synthetic* IV estimator for the regression of $\tilde{Y}$ on $\tilde{R}$ instrumented by $\tilde{Z}$ for $t > T_0$ recuperates the true parameter θ up to two potential bias terms.

$$\begin{aligned}
\tilde{\theta}^{TSLS} &= \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}\right)^{-1} \sum_{i,t>T_0} \tilde{Z}_{it}\tilde{Y}_{it} \\
&= \theta + \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}\right)^{-1} \sum_{i,t>T_0} \tilde{Z}_{it} \left(\mu_i - \sum_{j\neq i} \hat{w}_{ij}^{SC}\mu_j\right)' F_t \qquad (1.6) \\
&\quad + \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}\right)^{-1} \sum_{i,t>T_0} \tilde{Z}_{it} \left(\epsilon_{it} - \sum_{j\neq i} \hat{w}_{ij}^{SC}\epsilon_{jt}\right).
\end{aligned}$$

Similarly, the standard TSLS estimator $\hat{\theta}^{TSLS}$ can also be decomposed

$$\hat{\theta}^{TSLS} = \theta + \left(\sum_{i,t>T_0} Z_{it}R_{it}\right)^{-1} \sum_{i,t>T_0} Z_{it}\mu_i' F_t + \left(\sum_{i,t>T_0} Z_{it}R_{it}\right)^{-1} \sum_{i,t>T_0} Z_{it}\epsilon_{it}. \qquad (1.7)$$

In both cases, the bias in estimating θ will depend on a term involving the unobserved factor structure $\mu_i' F_t$ and a term involving the idiosyncratic error term ϵ_{it}. Under the partial instrument validity Assumption (2) we might expect the term involving $Z_{it}\epsilon_{it}$ to be close to zero in probability under suitable assumptions, but the term involving $\mu_i' F_t$ will cause omitted variable bias in the IV estimates. The intuition for the SIV estimator is that if the partialling out procedure successfully removes the $\mu_i' F_t$ term, thanks to the synthetic controls matching the factor loading μ_i for each unit, then there is no omitted variable bias. In section 1.4 we give conditions on the model primitives under which this is the case and the SIV estimator is consistent.

Common synthetic control weights The decompositions (1.5) and (1.6) clarify why it is necessary that a common set of synthetic control weights $\hat{w}_i^{SC}$ is used to generate the debiased outcomes $\tilde{Y}$ and treatments $\tilde{R}$. Suppose, that instead different weights w_1 and w_2 were used in constructing $\tilde{Y}_{it}$ and $\tilde{R}_{it}$. Let $\tilde{Y}_{it}^w$, $\tilde{R}_{it}^w$ and Z_{it}^w denote the debiased quantities in which weights w were used. In this case, the SIV estimator has the following decomposition

$$
\begin{aligned}
\tilde{\theta}_Z^{TSLS} &= \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}^{w^2}\right)^{-1} \sum_{it} \tilde{Z}_{it}\tilde{Y}_{it}^{w^1} \\
&= \theta \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}^{w^2}\right)^{-1} \sum_{it} \tilde{Z}_{it}\tilde{R}_{it}^{w^1} \\
&\quad + \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}^{w^2}\right)^{-1} \sum_{i,t>T_0} \tilde{Z}_{it} \left(\mu_i - \sum_{j\neq i} w_{ij}^1 \mu_j\right)' F_t \\
&\quad + \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}^{w^2}\right)^{-1} \sum_{i,t>T_0} \tilde{Z}_{it} \left(\epsilon_{it} - \sum_{j\neq i} w_{ij}^1 \epsilon_{jt}\right).
\end{aligned} \tag{1.8}
$$

The problem of using different weights is that the SIV estimator will estimate θ up to the term $G = \left(\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{R}_{it}^{w^2}\right)^{-1} \sum_{it} \tilde{Z}_{it}\tilde{R}_{it}^{w^1}$, which unless $w_1 = w_2$ may not be one in finite samples, and, in general, may not converge to one.

The focus of this paper is not to describe what parameters can be identified in this setting in terms of potential outcomes, as we consider the marginal effect defined in section 1.2. It is possible, however, to derive latent average treatment effect characterization in the case of discrete valued instruments and treatments under a modification of the standard monotonicity assumption. See Mogstad and Torgovitsky (2024) for an in-depth discussion of unobserved heterogeneity in treatment effect in IV models, and for discussion of identification in related IV difference-in-difference settings see Borusyak and Hull (2020). For a discussion of identification of continuous treatment effects in DiD designs we refer readers to de Chaisemartin et al. (2024).

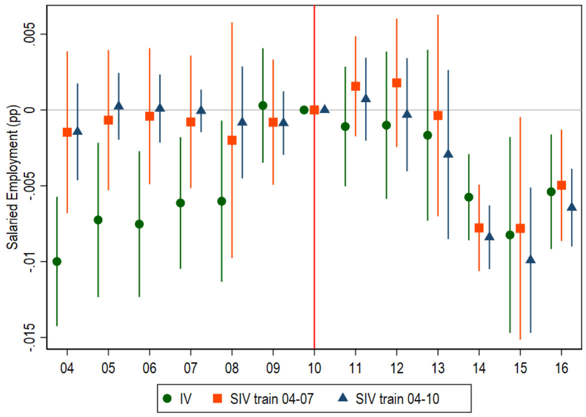
Example 1 (applying the SIV) We show how the synthetic IV estimator works by applying it to the Syrian refugee example. We proceed with the first step by computing the synthetic control weights for each Turkish region by solving problem (1.3) with $\mathcal{W} = \{w \mid w_j \geq 0, \quad \sum_j w_j = 1\}$.[5] We then compute the debiased variables $\{\tilde{Y}_{it}, \tilde{R}_{it}, \tilde{Z}_{it}\}_{t=T_0+1}^T$ and estimate the reduced form regression (1.1) with the debiased data

$$
\tilde{Y}_{it} = \sum_{k\neq 2010} \tilde{\theta}_k(\mathbb{1}\{t = k\} \times \tilde{Z}_i) + \alpha_i + \delta_t + \epsilon_{it}, \tag{1.9}
$$

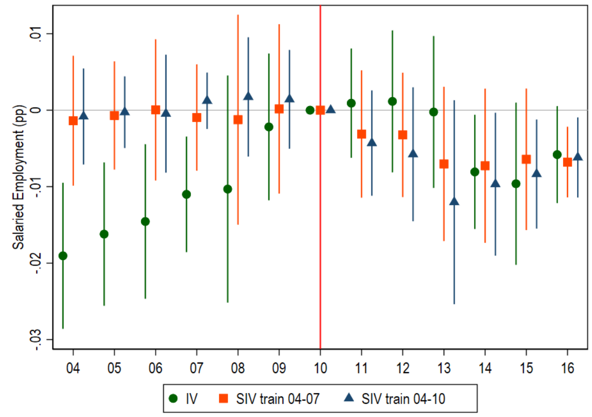
where $\tilde{Z}_i$ is the debiased instrument share component. In panel(a) of Figure 1.4 we plot the event study estimates for salaried employment as we did in Figure 1.2. As we saw before, the

[5]As an additional step, we normalize the outcome variable Y before solving for the weights.

Figure 1.4: Comparing IV and SIV for the Syrian refugee shock.



(a) Salaried employment.



(b) Salaried employment for men.

Notes: panel (a) shows the event-study estimates for regressions 1.1 and 1.9 for the shift-share IV and SIV for salaried employment respectively; it replicates panel (d) in Figure 1.2 for the SIV. Panel (b) shows the same event-study estimates for salaried employment for men. Two SIV estimators are estimated, SIV train 04-10 uses all the pre-treatment periods to compute the synthetic weights in program (1.3), while SIV train 04-07 only uses the first 4 time periods.

IV estimator (green circles) exhibits large pre-trends, but the SIV estimator (blue triangles) does not. To check that indeed the absence of pre-trends for the SIV estimator is not due to over-fitting to the pre-treatment period idiosyncratic noise (ϵ_{it}), we also estimate a backdated SIV in which only a subset of the pre-treatment periods is used in estimating the weights in program (1.3). The backdated SIV (in orange squares) also shows no pre-trends, providing evidence that the synthetic IV estimator is successfully capturing the unobserved U_{it} term. In the following sections, we highlight the theoretical properties of the SIV estimator and other empirical checks, such backdating, researchers can do to check that robustness of the estimator. In the Appendix (section A.1.7), we extend the theoretical results for the SIV estimator to the event study design considered in this example and the empirical applications.

A feature of the synthetic IV estimator is that it accounts for unmeasured confounding that may change over time, and, crucially, differentially before and after the intervention at T_0. This can be seen in Figure 1.4. In panel (a), while the SIV accounts for the pre-trends before T_0, the estimates after T_0 see changes smaller in magnitude relative to the pre-trend size and non-linear in time, with effects increasing at the start of the post-treatment period and decreasing at the end of the period. If researchers had estimated a linear trend instead ($U_{it} = \alpha_i \times t$) as is common in the literature, the post-treatment estimates would have been shifted downwards significantly. The flexibility of the SIV allows for cases in which the trend in the post-treatment period might be different, for example if we believe Turkish regions close to Syria are in growing path (catching up to richer regions) we might expect the confounding to be smaller in the post-treatment period. Furthermore, the SIV also allows flexibility in functional form across outcomes. In panel (b) of Figure 1.4 we show the event-study estimates for a different outcome, formal salaried employment for men. While the pre-trends for this outcome are similar than for salaried employment, the SIV post-treatment estimates are shifted downwards uniformly. In section 1.7.1 we revisit this empirical example and show that SIV can make a real difference compared to the standard IV in estimating the effects of refugees on several local labor outcomes. To preview the results, a researcher using IV would find no effect on natives' or men's salaried employment. Using SIV, however, we find a statistically significant negative effect for both outcomes. We discuss the relevance of these findings in section 1.7.1.

In sections 1.4 and 1.6 we discuss the theoretical properties of the synthetic IV estimator and investigate its finite sample properties through simulations. We highlight that in well behaved settings with low noise but significant correlation between the instrument and the unobserved factor structure the SIV estimator can provide reliable estimates of the true effect. In section 1.7 we re-visit our three empirical applications using the SIV estimator.

1.4 Theoretical Results

In this section we provide theoretical guarantees for the SIV estimator. To characterize the behavior of the standard IV and the synthetic IV estimators it is key to understand the behavior of the terms involving the unobserved factor μ_i in decompositions (1.6) and (1.7). In order to do so, we impose more structure on the primitives of the design described in assumptions 1 and 2.

Assumption 3 (Model primitives)**.** *Assumptions on the factor structure, the error components and the instruments are as follows.*

- *The common factors are bounded such that for all t, $|F_{lt}| \leq \bar{F}$ for $l = 1, \ldots, k$. Furthermore, the matrix $F_{T_0}F'_{T_0}$ has minimum eigenvalue ξ such that $\xi/T_0 > 0$, where F_{T_0} is the $k \times T_0$ matrix of common factors F_t for $t \leq T_0$. The factor loadings are bounded such that for all i $|\mu_i| \leq c_\mu$.*
- *The unobserved term A_{it} is bounded such that for all i, t $|A_{it}| \leq c_A$. Furthermore, $\frac{1}{JT_1}\sum_{i,t>T_0} A_{it} \xrightarrow{p} 0$ as $JT_1 \to \infty$.*
- *The instrument $Z_{it} \in \mathcal{Z}$ is bounded such that for all i, t $|Z_{it}| \leq c_z$ and $\frac{1}{JT_1}\sum_{i,t>T_0} Z_{it}^2 \xrightarrow{p} Q_Z > 0$.*

- *The instrument Z_{it} and the unobserved factor structure satisfy*

$$\frac{1}{JT_1} Z' M_U Z \xrightarrow{p} Q > 0,$$

as $JT_1 \to \infty$ for $T_1 = T - T_0$, where $Z = vec(Z_1^T)$ is a $JT_1 \times 1$ vector of instruments and $M_U = I - U_{JT_1}(U'_{JT_1} U_{JT_1})^{-1} U'_{JT_1}$ is the $JT_1 \times JT_1$ residual maker matrix for $U_{JT_1} = vec(U^{T_1})$. Furthermore, the first stage parameter satisfies $\gamma > 0$.

- *ϵ_{it} and η_{it} are i.i.d mean zero subGaussian random variables with variance σ_ϵ^2 and σ_η^2 respectively, finite covariance $\sigma_{\epsilon\eta} = \mathbb{E}[\epsilon_{it}\eta_{it}]$ and bounded fourth moments.*

Assumption 3 has three parts. First, we assume that the model primitives are bounded. This is a common assumption in papers analyzing the behavior of synthetic control estimators and rules out weak factors. Second, we assume that the instrument is strong and not perfectly correlated with the unobserved factor structure. That is, after projecting out the unobserved confounder U_{it} enough variation remains in the instrument. This requirement avoids weak instrument problems and in the simulation discussion we highlight the importance of this assumption for the finite sample performance of the synthetic IV estimator. Finally, we assume that the unobserved error terms η and ϵ are i.i.d, but potentially correlated. This assumption can be weakened to allow for time series correlation, however in our main results the time series dependence is present through the unobserved factor structure $\mu_i' F_t$.

Observe that under our design and Assumption 3 the term in decomposition (1.7) and (1.6) depending on the idiosyncratic shocks will converge to zero in probability. In the appendix, we derive a finite sample bound for this term (see Lemma 2). On the other hand, the unobserved factor term $\sum_{i,t>T_0} Z_{it}\mu_i' F_t$ need not converge to zero in probability at a $1/(JT_1)$ rate. Therefore, in general the TSLS estimator, with or without fixed effects, will be asymptotically biased. The synthetic IV estimator will also be biased in finite samples as $\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{\mu}_i' F_t$ need not be zero, but the bias will depend on how well the synthetic control procedure in step 1 can partial out the factor loadings $\tilde{\mu}_i$.[6] To give conditions for consistency, we condition on the unobserved factor structures and consider ϵ, η and the instrument Z as the source of randomness in our design. The following result gives a bound on the $\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{\mu}_i' F_t$ in terms of model primitives and the mean-absolute deviation of pre-treatment values of Y and R, defined as $\text{MAD}(\tilde{Y}^{T_0}) = \frac{1}{JT_0}\sum_{i,t\leq T_0} |\tilde{Y}_{jt}|$.

[6] Under weak conditions on the time series (β-mixing, or covariance stationarity), it may also be possible to directly correct this bias using a cross-fitting procedure in the spirit of the method proposed in Chernozhukov et al. (2022) for the synthetic control framework.

Theorem 1 (Factor term bound)**.** *Under Assumptions 1-3, for $t > T_0$ conditional on the unobserved components $\mu_i' F_t$ and A_{it}, the following bound holds for all J, T_1 and T_0*

$$\frac{1}{JT_1}\mathbb{E}\left[\left|\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{\mu}_i' F_t\right|\right] \leq \left(\frac{\bar{F}^2 k c_z c}{\xi}\right)\left(2c\sqrt{\frac{J}{T_0}}\sigma_\epsilon + \mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}_{jt}|\right] + \theta\mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{R}_{jt}|\right]\right)$$

where $c = 1 + C$ and all other terms are defined in Assumptions 1-3.

Theorem 1 states that the bias term that depends on the unobserved factor structure can be bounded above by the expected mean absolute deviation of the outcome variable in the pre-treatment period and a term that depends on the likelihood of pre-treatment “over-fitting". This is a standard bound in papers evaluating the properties of synthetic control estimators (see Abadie et al. (2010) for the first example in the literature and Vives-i-Bastida (2022) for a example with covariates). It highlights the dependence of the estimator on good pre-treatment fit (see Ferman and Pinto (2021) for a discussion of synthetic controls with imperfect pre-treatment fit). In particular, the bound depends on the error noise level σ_ϵ and the ratio $\sqrt{J/T_0}$. In settings, in which the we have a small amount of pre-treatment periods, a large number of units, or in which the noise level is high, perfect interpolation of the noise is more likely, biasing the estimator. A discussion in Abadie and Vives-i-Bastida (2025) highlights the importance of pre-treatment fit and over-fitting for performance of synthetic control estimators through a simulation study. Similarly, we evaluate the performance of the synthetic IV estimator in simulations in section 1.6 and find that the estimator performs well even in settings with moderate $\sigma_\epsilon\sqrt{J/T_0}$.

To provide conditions under which the estimator is consistent as JT_1 grows to infinity, we consider a relaxation of rank proposed by Rudelson and Vershynin (2007) that allows for *small* perturbations.

Assumption 4 (Numerical rank assumption)**.**

With probability one, for all J and T_0, the pre-treatment matrices have bounded numerical rank, $\frac{\|Y^{T_0}\|_F^2}{\|Y^{T_0}\|_2^2} \leq \bar{r}_1$ $\frac{\|R^{T_0}\|_F^2}{\|R^{T_0}\|_2^2} \leq \bar{r}_2$, and their largest singular values are bounded above such that $\sigma_1(Y^{T_0}) \leq \bar{\sigma}_1$ and $\sigma_1(R^{T_0}) \leq \bar{\sigma}_2$, where $\bar{r}_1$, $\bar{r}_2$, $\bar{\sigma}_1$ and $\bar{\sigma}_2$ may depend on J and T_0.

The intuition behind Assumption 4 is better seen by considering the rank of the $J \times T_0$ design matrix Y^{T_0}. If the matrix had fixed rank $r < \min\{T_0, J\}$ all points would lie in a low dimensional manifold of the space and the pre-treatment fit error would grow proportional to r. Given that in our setting the error terms are *i.i.d* shocks, this is not a reasonable assumption. Instead, we consider a bound on the numerical rank; the ratio between the Frobenius and 2-norm of a matrix. This notion of rank allows for points to lie “close" to a low dimensional manifold. Furthermore, for any matrix A it follows that

$$\frac{\|A\|_F^2}{\|A\|_2^2} \leq \text{rank}(A),$$

therefore the bounded numerical rank assumption is implied by a bounded rank assumption. Whether Assumption 4 is satisfied will depend on the model primitives. In particular, it will be satisfied when the signal to noise ratio is high. That is, when the factor structure $\mu_i' F_t$ dominates the noise term ϵ. In cases in which σ_ϵ is large relative to the factor term the numerical rank is likely to be large and the pre-treatment fit bad. In section 1.6 we explore the performance of our estimator in a variety of settings and propose checks researchers can implement to evaluate whether their empirical setting is likely to satisfy this assumption.

Theorem 2 (Factor term consistency). *Under Assumptions 1-4, for $t > T_0$ conditional on the unobserved components $\mu_i' F_t$ and A_{it}, the following bound holds for all J, T_1 and T_0*

$$\frac{1}{JT_1}\mathbb{E}\left[\left|\sum_{it} \tilde{Z}_{it}\tilde{\mu}_i' F_t\right|\right] \leq \left(\frac{\bar{F}^2 k c_z c}{\underline{\xi}}\right)\left(2c\sqrt{\frac{J}{T_0}}\sigma_\epsilon + (\sqrt{\bar{r}_1\bar{\sigma}_1} + \theta\sqrt{\bar{r}_2\bar{\sigma}_2})\left[\frac{1}{\sqrt{JT_0}} + C\sqrt{\frac{1}{T_0}}\right]\right)$$

where $c = 1 + C$ and all other terms are defined in the assumptions. Furthermore, as $JT_1 \to \infty$, $(\sqrt{\bar{r}_1\bar{\sigma}_1} + \theta\sqrt{\bar{r}_2\bar{\sigma}_2})\sqrt{\frac{1}{T_0}} \to 0$, and $\sqrt{\frac{J}{T_0}} \to 0$,

$$\frac{1}{JT_1}\sum_{it} \tilde{Z}_{it}\tilde{\mu}_i' F_t \xrightarrow{p} 0.$$

Theorem 2 shows that the bias due to the factor term is $o_p(1)$ as long as $(\bar{r}_1\bar{\sigma}_1 + \bar{r}_2\bar{\sigma}_2)\frac{1}{T_0} \to 0$. For fixed J, this implies that we need $T_0, T_1 \to \infty$. The restrictions on the rank and $\bar{\sigma}_1$, $\bar{\sigma}_2$ are not uncommon in the matrix completion literature. Combining the consistency result with the additional assumptions on the instrument behavior we can show that the synthetic IV estimator is a consistent estimator of θ.

Theorem 3 (Consistency). *Under Assumptions 1-4, as $JT_1 \to \infty$, $(\bar{r}_1\bar{\sigma}_1 + \theta\bar{r}_2\bar{\sigma}_2)\frac{1}{T_0} \to 0$ and $\frac{J}{T_0} \to 0$,*

$$\tilde{\theta}^{TSLS} - \theta \xrightarrow{p} 0,$$
$$\tilde{\theta}_Z^{TSLS} - \theta \xrightarrow{p} 0,$$
$$\tilde{\theta}_{YR}^{TSLS} - \theta \xrightarrow{p} 0.$$

Theorem 3 states that both the synthetic IV estimator $\tilde{\theta}^{TSLS}$ and the synthetic IV estimators for which we do not debias the instrument $\tilde{\theta}^{TSLS}_{YR}$ and for which we only debias the instrument $\tilde{\theta}^{TSLS}_{Z}$, are consistent given our assumptions and the rate conditions of Theorem 2. Under our model and assumptions, the standard TSLS estimator will not be consistent in general as $\frac{1}{JT_1}\sum_{it} Z_{it}\mu_i' F_t$ may not converge in probability to zero. As discussed, however, the synthetic IV estimator is biased in finite samples and the finite sample bias will depend on the signal to noise ratio, the length of the pre and post treatment periods in relation to the number of units J and, through the first stage, the correlation between Z_{it} and $\mu_i' F_t$. It is important to note that while debiasing the instrument does not affect the consistency of the estimator it may improve the finite sample properties of the estimator. In the appendix, we show under additional assumptions, that debiasing the instrument can lead to a stronger first stage and lower correlation with the debiased unobserved term $\tilde{\mu}_i' F_t$, leading to better finite sample properties. We confirm this intuition in the simulation study by comparing $\tilde{\theta}^{TSLS}$ and $\tilde{\theta}^{TSLS}_{Z}$. Finally, under Assumptions 1-4 it is also possible to show that the synthetic IV estimator is asymptotically normal.

Theorem 4 (Asymptotic normality)**.** *Under Assumption 1-4, conditional on weights w and instruments Z_{it}, if $\frac{T_1}{T_0}(1+J)(\bar{r}_1\bar{\sigma}_1 + \theta\bar{r}_2\bar{\sigma}_2) \to 0$ and $\frac{1}{\sqrt{JT_1}}\max_i \sum_{j\neq i}|w_{ji}| \to 0$, then as $JT_1 \to \infty$*

$$\frac{\sqrt{JT_1}(\tilde{\theta}^{TSLS} - \theta)}{v_{JT_1}} \xrightarrow{d} (\gamma Q)^{-1} \times N(0,1).$$

where $v^2_{JT_1} = \frac{1}{JT_1}\sum_{it} var(\tilde{Z}_{it}\tilde{\epsilon}_{it} \mid Z, w) = \frac{1}{JT_1}\sum_{i,t>T_0}\sigma^2_\epsilon \tilde{\alpha}^2_{it}$ and $\tilde{\alpha}_{it} = \tilde{Z}_{it} - \sum_{j\neq i}\tilde{Z}_{jt}w_{ji}$. Furthermore, given $w_i \in \mathcal{W}$, a sufficient condition for $\frac{1}{\sqrt{JT_1}}\max_i \sum_{j\neq i}|w_{ji}| \to 0$ is that $\frac{J}{T_1} \to 0$ as $JT_1 \to \infty$.

Theorem 4 shows that the estimator converges to a normal random variable centered at the true parameter when normalized by the conditional variance v_{JT_1} which depends on the instruments Z and synthetic control weights w. The result allows us to construct standard asymptotic confidence intervals by using the sample counterparts. Let $\tilde{\sigma}^2_{TSLS}$ denote the variance of the synthetic TSLS estimator which is given by

$$\tilde{\sigma}^2_{TSLS} = \frac{JT_1\hat{v}^2_{JT_1}}{(\sum_{i,t>T_0}\tilde{Z}_{it>T_0}\tilde{R}_{it})^2} = \frac{\hat{\sigma}^2_\epsilon\|\tilde{\alpha}\|^2_2}{(\sum_{i,t>T_0}\tilde{Z}_{it}\tilde{R}_{it})^2},$$

where $\tilde{\sigma}^2_\epsilon$ can be estimated from the regression residuals, and the denominator and $\|\tilde{\alpha}\|^2_2 = \sum_{it}\tilde{\alpha}^2_{it}$ can be computed directly from the data. After computing this quantity, standard

$(1-\alpha)\%$ confidence intervals can be constructed such that

$$\theta \in \left[\tilde{\theta}_{TSLS} - z_{1-\alpha/2} \times \frac{\tilde{\sigma}_{TSLS}}{\sqrt{JT_1}}, \tilde{\theta}_{TSLS} + z_{1-\alpha/2} \times \frac{\tilde{\sigma}_{TSLS}}{\sqrt{JT_1}}\right], \tag{1.10}$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution.

The result in Theorem 4 requires an additional density condition with respect to the conditions for consistency in Theorem 2. The condition requires that for the sequence of synthetic control weights w, $\frac{1}{\sqrt{JT_1}} \max_i \sum_{j \neq i} |w_{ji}| \to 0$. Intuitively, it ensures that the weights are not concentrated on a few units such that, as JT_1 grows, the estimator does not depend on a few data points. While, in general, whether this condition is satisfied will depend on the model primitives, under the l_1-norm constraint in program (1.3), a sufficient condition is that $J/T_1 \to 0$ as $JT_1 \to \infty$. In finite samples, however, we can inspect directly whether our estimated weights $\hat{w}$ are dense or not. In the simulation exercise and empirical applications we show that in general the weights are dense, explaining the good behavior of the SIV estimator in shorter panels.

Variance comparison. A natural question is how does the variance of the SIV estimator compare with the variance of the standard TSLS. Which variance is larger will depend on the model primitives and the relationship between the unobserved confounder U and the instrument Z. The intuition is similar to that of the effect of adding additional covariates on the variance of the OLS estimator for a parameter of interest in a linear model. Adding additional covariates reduces the variance of the unobserved component, but this may come at the cost of variance inflation due to the correlation between the regressor of interest and the additional covariates.

In our setting, if the SIV estimator is successful in partialling out the unmeasured confounder U, and the variance of U is large relative to σ_ϵ^2 or the correlation between U and Z is sufficiently small, then the SIV estimator will exhibit a smaller variance than the standard TSLS. To see this, suppose that in addition to Assumptions 1-4, $U_{it} = \mu_i' F_t$ are i.i.d random variables and $var(\mu_i' F_t) = \sigma_u^2 > 0$. Then, it follows that the asymptotic variances of the TSLS and SIV estimators are given by $\sigma_{TSLS}^2 = (\gamma^2 Q_Z)^{-1}[\sigma_u^2 + \sigma_\epsilon^2]$ and $\sigma_{SIV}^2 = (\gamma^2 Q)^{-1}\sigma_\epsilon^2$. Therefore, the ratio of the variances is greater than one when

$$\frac{\sigma_{TSLS}^2}{\sigma_{SIV}^2} = \frac{\sigma_u^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2} \frac{Q}{Q_Z} > 1 \iff \frac{\sigma_u^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2} > \frac{Q_Z}{Q}. \tag{1.11}$$

The LHS in the last expression in (1.11) is the signal to noise ratio, which given that $\sigma_u^2, \sigma_\epsilon^2 > 0$, is strictly greater than one. The RHS depends on the correlation between Z and U. Recall that Q is the residual variation in Z after U has been projected out, therefore $Q_Z/Q \geq 1$, with $Q_Z/Q = 1$ when $corr(Z,U) = 0$ and $Q_Z/Q \to \infty$ when $corr(Z,U) \to 1$ and $Q \to 0$. It follows that when $corr(Z,U) = 0$, the SIV estimator will have strictly lower variance than the TSLS, and when $corr(Z,U) \to 1$ the SIV will have greater variance than the TSLS. Intuitively, when $corr(Z,U) = 0$ this is exactly the case in which controlling for additional covariates would lower the variance of the OLS estimator of θ when R is randomly assigned (e.g. in an RCT). As the correlation between Z and U increases which variance dominates will depend on the signal to noise ratio. We investigate this trade off in the simulation exercise and the empirical applications and find that often the SIV exhibits lower variance than the TSLS.

1.5 Extensions

1.5.1 Combining SIV with additional estimators

The set up described under Assumptions 1 and 2 highlights a trade off between using the instrument variation to address the endogeneity bias due to the correlation between ϵ and η and incurring an omitted variable bias due to the instrument's correlation with the unobserved term $\mu_i' F_t$. The synthetic IV estimator can address these biases asymptotically in regimes in which σ_ϵ is small relative to the variation in $\mu_i' F_t$ as we highlighted in section 1.4. However, when σ_ϵ is large the endogeneity concern becomes more important than the omitted variable bias and, therefore, we might be able to design an estimator that addresses this bias more directly. With this in mind, we consider an additional estimator that will perform better in cases in which the noise level is high, and propose an ensemble estimator as a 'doubly robust' alternative to the synthetic IV.

Suppose that the instrument also follows a factor structure, such that $Z_{it} = Z_i' G_t$ for factor loadings Z_i and common factors G_t. This is the case in shift share designs such as the Syrian refugee example or the China shock study. In such cases, a natural estimator robust to noise ϵ_{it} is one that computes the synthetic control weights after projecting the outcome variable in the instrument space. The intuition for this estimator is that the outcome Y_{it} is noisy due to the unobserved error ϵ, but given our partial instrument validity assumption 2, after projecting the outcome in the instrument space we partial out the noise.

The *projected* synthetic estimator can be computed similarly to the SIV estimator, with an additional step.

1. Project to instrument space: for $t \leq T_0$, let $Y_{zt} = Z(Z'Z)^{-1}Z'Y_t$, where $Z = (Z_1, \ldots, Z_J)'$ and $Y_t = (Y_{1t}, \ldots, Y_{Jt})'$ are $J \times 1$ vectors.

2. Use the de-noised outcomes to compute the SC weights in the pre-period

$$\hat{w}_j^P \in \text{argmin}_{w \in \mathcal{W}} \|Y_j^{T_0} - Y_{z,-j}^{T_0}{}' w\|^2.$$

3. Define the de-biased quantities $\tilde{Y}_{it}^P$, $\tilde{Z}_{it}^P$, $\tilde{R}_{it}^P$ accordingly for the projected weights $\hat{w}_j^P$.

4. For the post treatment period $t > T_0$, estimate the synthetic TSLS projected estimator

$$\tilde{\theta}^P = \left(\sum_{it} \tilde{Z}_{it}^P \tilde{R}_{it}^P\right)^{-1} \sum_{it} \tilde{Z}_{it}^P \tilde{Y}_{it}^P.$$

The performance of the projected estimator vis-a-vis the SIV will depend on the data generating process primitives. In cases in which the noise error term ϵ_{it} is more important than the factor term $\mu_i' F_t$ the projected estimator will perform favorably. On the other hand, if the factor structure (the signal) dominates, the projected estimator will perform worse than the synthetic IV as it will fit the factor structure $\mu_i' F_t$ worse. To see this, consider the bound in Theorem 1 for the projected estimator. In the appendix we show that with high probability under Assumptions 1-4

$$\frac{1}{JT_1}\mathbb{E}\left[\left|\sum_{i,t>T_0} \tilde{Z}_{it}^P \tilde{\mu}^{P_i\prime} F_t\right|\right] \leq \left(\frac{\bar{F}^2 k c_z c}{\xi}\right)\left(2c\sqrt{\frac{1}{T_0}}\sigma_\epsilon + \mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0} |\tilde{Y}_{jt}^P|\right] + \theta\mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0} |\tilde{R}_{jt}^P|\right]\right),$$

where all the terms are defined in Assumption 3 and $c = 1 + C$. This bound differs from the bound in Theorem 1 in two important ways. First, the contribution of the noise term ϵ_{it}^2 to the overfitting bias changes from scaling with $\sqrt{\frac{J}{T_0}}$ to scaling with $\frac{1}{\sqrt{T_0}}$. This is because the estimated weights $\hat{w}_j^P$ dependence on the idiosyncratic shock ϵ_{it} vanishes asymptotically given that we are projecting into the instrument space and $Z_i \perp \epsilon_{it}$. Second, the pre-treatment fit will be worse as the weights only use variation in the instrument space. Therefore, while the projected estimator will be consistent under the assumptions and asymptotic regime of Theorem 2, its finite sample properties will differ from those of the SIV estimator. In particular, we expect the projected estimator to perform better than the SIV when

$$0 \leq \Delta_{T_0}^P - \Delta_{T_0}^{SIV} \leq \sqrt{J}\left(\frac{2(1+C)\sigma_\epsilon}{\sqrt{T_0}}\right), \tag{1.12}$$

where for an estimator $a \in \{\text{SIV}, \text{P}\}$ the expected pre-treatment fit is given by $\Delta_{T_0}^a =$

$\mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}^a_{jt}|\right]+\theta\mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{R}^a_{jt}|\right]$. Which means that in settings in which the noise level is large (high σ_ϵ) or the relative of number of units to pre-treatment periods is not close to zero, the projected estimator may have smaller finite sample bias than the SIV. While the bias itself cannot be evaluated directly in practice, by comparing differences in pre-treatment fit (LHS in 1.12) and checking the noise level of a given empirical setting (to evaluate the RHS in 1.12), we can have an idea of which estimator might be more or less biased.

Given that each estimator has a different finite sample bound, that depends on different model primitives, we can construct an ensemble estimator that combines both estimators and is robust to different sources of bias. For a hyper-parameter $\alpha^h \in [0,1]$ we define the ensemble estimator as

$$\tilde{\theta}^E(\alpha^h) = \alpha^h\tilde{\theta}^{TSLS} + (1-\alpha^h)\tilde{\theta}^P,$$

The α hyper-parameter can be chosen through cross-validation in the pre-period to optimize the mean squared error of the synthetic control estimator. The following steps detail how to compute the ensemble estimator.

1. Split the pre-period into a training period $1,\ldots,T_v$ and a validation period $T_v+1,\ldots,T_0$.

2. In the training period compute the synthetic control weights for each estimator, $\hat{w}^P$ and $\hat{w}$, and the debiased outcomes $\tilde{Y}^P_{it}$ and $\tilde{Y}_{it}$.

3. In the validation period choose α^* to minimize the mean squared error in the validation

$$\frac{1}{J(T_0-T_v)}\|\alpha^h\tilde{Y}^{P,T_v}+(1-\alpha^h)\tilde{Y}^{T_v}\|_2^2,$$

   where $\tilde{Y}^{T_v}$ denotes the debiased outcomes for the validation period.

4. Compute the ensemble estimator in the post period as $\alpha^*\tilde{\theta}^{TSLS}+(1-\alpha^*)\tilde{\theta}^P$.

In the appendix, we show that $\tilde{\theta}^E(\alpha^h)$ is a consistent estimator of θ for any $\alpha^h \in [0,1]$, and therefore, the cross-validated ensemble estimator is also consistent. The finite sample improvement of the cross-validated estimator will, however, depend on the finite sample bias differences between the estimators and the length of the validation period used to calibrate the hyper-parameter. In the simulation exercise in section 1.6, we show that the projected estimator performs well, with slightly worse performance to the SIV in low noise cases and better performance in high noise cases. In the Appendix (section A.1.8) we also propose an alternative estimator based on a time-series aggregation scheme.

1.5.2 Alternative inference procedures

The asymptotic results in section 1.4 require that $JT_1 \to \infty$ and $J/T_0 \to 0$. However, in many shift-share IV and synthetic control design settings the researchers may have at their disposal a moderate number of units and time periods. This is the case of our main empirical application to the Syrian civil war. With this in mind, the literature has considered permutation based tests (Abadie et al. (2010), Abadie and Zhao (2022), Firpo and Possebom (2018)) and randomization inference procedures (Imbens and Rosenbaum (2005), Borusyak and Hull (2020)) as alternatives to asymptotic based confidence intervals. In this section, we describe how a split conformal inference procedure can be applied in the context of the synthetic IV. In the appendix, we describe an alternative randomization inference procedure.

Split conformal inference Our discussion in section 1.2 high-lights the use of reduce-form event studies as a way to assess if there is unmeasured confounding in an instrument $Z_{it} = Z_i \times \bar{H}_t$. In the Appendix (section A.1.7) we expand our main theoretical set up to include the event study designs. The event studies take the following form

$$Y_{it} = \sum_{k \neq T_0} \theta_k (\mathbb{1}\{t = k\} \times Z_i) + \epsilon_{it}, \tag{1.13}$$

where, because $Z_{it} = 0$ for $t \leq T_0$, in absence of unmeasured confounding, we have that $\theta_l = 0$ for $l \leq T_0$. Therefore, in our linear IV design (1), testing the null $H_0 : \theta = 0$ is implied by testing that $\{\theta_l = \theta_k$ for all $l \leq T_0$ and $k > T_0\}$. We propose a permutation based test for this null following the split-conformal inference procedures of Abadie and Zhao (2022) and Chernozhukov et al. (2021). The test can be implemented with any estimator of the event study coefficients θ_k, but we detail the procedure for the SIV estimator.

1. Split $1, \ldots, T_0$ into a *training* period $1, \ldots, T_b$ and a *blank* period $T_b + 1, \ldots, T_0$.

2. Compute SC weights in the training period and define debiased quantities accordingly.

3. Run reduced form event regression as in (1.13) using the debiased quantities $\tilde{Y}_{it}$ and $\tilde{Z}_i$ to get estimates $\{\tilde{\theta}_{T_b+1}, \ldots, \tilde{\theta}_T\}$.

4. Generate $T_1 \times 1$ permutation vectors $\theta_\pi = (\tilde{\theta}_{\pi(1)}, \ldots, \tilde{\theta}_{\pi(T_1)})$ for $\pi \in \Pi$, where Π is the set of size T_1 combinations from $T_b, \ldots, T$.

5. Compute the permutation test statistic $S(\theta) = 1/(T - T_0)\|\theta\|_1$ for each $\pi \in \Pi$.

6. Compute the permutation p-value:

$$\hat{p} = \frac{1}{\Pi} \sum_{\pi \in |\Pi|} \mathbf{1}(S(\tilde{\theta}_\pi) \geq S(\tilde{\theta}_{t>T_0}))$$

Under our design Assumptions 1-3, if additionally, $Z_{it} = Z_i \times \bar{H}_t$ and $\{F_t\}_{t>T_b}$ is a sequence of exchangeable random variables independent of ϵ_{it} and η_{it}, it follows that tests based on $\hat{p}$ are exact in the sense that under the null H_0, for $\alpha \in [0,1]$ we have that

$$\alpha - \frac{1}{|\Pi|} \leq P(\hat{p} \leq \alpha) \leq \alpha$$

where P is taken over the distribution of $\{\epsilon_{it}, F_t, \eta_{it}\}$. Note that given that Assumption 3 ensures that ϵ_{it} and η_{it} are $i.i.d$, it follows that for all i, $\{\epsilon_{it}, F_t, \eta_{it}\}$ are exchangeable in t and the test exact validity result follows directly from Chernozhukov et al. (2021). For a result that relaxes the exchangeability assumption we refer readers to Abadie and Zhao (2022).

The permutation based inference procedure is going to be exactly valid in settings in which the time series structure of the unobserved confounder satisfies the exchangeability restriction. This is in contrast to the CI based on the SIV variance estimator (1.10) which are valid asymptotically under the regime of Theorem 4. In general, however, the power of the permutation based tests may be smaller and will depend on the number of time periods available and the noise levels. In the simulations, in section 1.6, we highlight the complementarity of both inference procedures in detecting the true effects using the SIV estimator.

1.6 Simulation study

In this section, we consider a simulation design calibrated to the Syrian empirical application. In the appendix, we consider different simulation designs with varying number of units, time periods, instrument strength, signal-to-noise ratios and number of factors. All simulation designs follow the following data generating process

$$\begin{aligned} Y_{it} &= \theta R_{it} + \mu_i' f_t + \epsilon_{it}, \\ R_{it} &= (\gamma Z_{it} + \eta_{it}) * \mathbb{1}(t \geq T_0), \\ Z_{it} &= Z_i' g_t * \mathbb{1}(t \geq T_0), \end{aligned}$$

with time series structure

$$f_t = \kappa_f f_{t-1} + u_{ft},$$
$$g_t = \kappa_g g_{t-1} + u_{gt},$$

and error structure

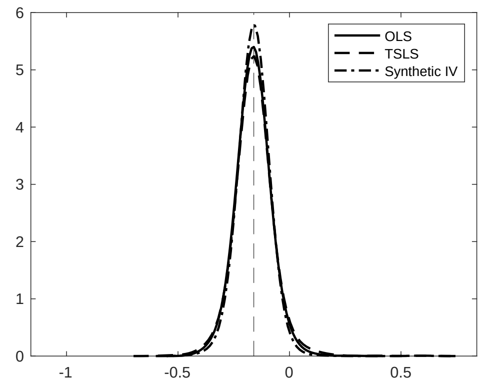
$$\begin{pmatrix} u_{ft} \\ u_{gt} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_f^2 & \rho_g \sigma_f \sigma_g \\ \rho_g \sigma_f \sigma_g & \sigma_g^2 \end{bmatrix}\right),$$
$$\begin{pmatrix} Z_i \\ \mu_i \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_z^2 & \rho_z \sigma_z \sigma_\mu \\ \rho_z \sigma_z \sigma_\mu & \sigma_\mu^2 \end{bmatrix}\right),$$
$$\begin{pmatrix} \epsilon_{it} \\ \eta_{it} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_\epsilon^2 & \rho \sigma_\epsilon \sigma_\eta \\ \rho \sigma_\epsilon \sigma_\eta & \sigma_\eta^2 \end{bmatrix}\right).$$

To map our simulation study to the data, we set the number of time periods to $T = 16$, with the intervention at $T_0 = 10$ and consider $J = 26$ regions. We target a relatively small true parameter of -0.16, with $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.035$ calibrated to the residual variance in the data (noting that for some outcomes the variance is significantly smaller) and set σ_Z^2 and γ such that the F-statistic is 150. Finally, we let the signal be given by $\sigma_\mu^2 = 0.25$, consider one factor $k = 1$ (as it explains 98% of the variance in Y from PCA) and let the AR parameter be $\kappa = \kappa_f = \kappa_g = 0.5$. Given this design, we proceed by varying ρ, ρ_z, ρ_g, which given σ_Z^2 changes the correlation structure between the unobserved terms and the instrument (Q/Q_Z) and σ_ϵ, which given σ_μ changes the signal-to-noise ratio $\left(\frac{\sigma_\mu^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2}\right)$ to evaluate the performance of the estimators in different settings.

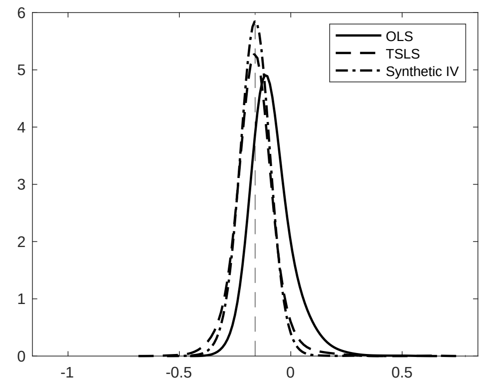
Figure 1.5 shows that the synthetic IV estimator is able to correct the bias present in the OLS and TSLS (with two-way fixed effects) estimators when there is endogeneity and omitted variable bias. Panel (a) shows the case in which all estimators are consistent (no correlation between R_{it} and U_{it} or ϵ_{it}). In this base case, as expected, the synthetic IV estimator performs similar to the OLS and TSLS estimators. In panel (b) we increase the correlation between ϵ_{it} and η_{it}, creating an endogeneity problem that can be addressed using the instrument. The OLS estimator is now biased, while the TSLS and synthetic IV estimators remain unbiased.[7] In panel (c) we introduce correlation between the instrument and the unobserved factor structure, by setting $\rho = \rho_z = \rho_g = 0.5$, and the instrument becomes invalid leading to biased TSLS estimates despite adding two-way fixed effects in the specification. The synthetic IV

[7]In this design the bias due to the correlation between ϵ_{it} and η_{it} is small given that their variances are small relative to σ_μ and σ_z. In the simulation table and in the appendix we consider designs in which this bias is more important.

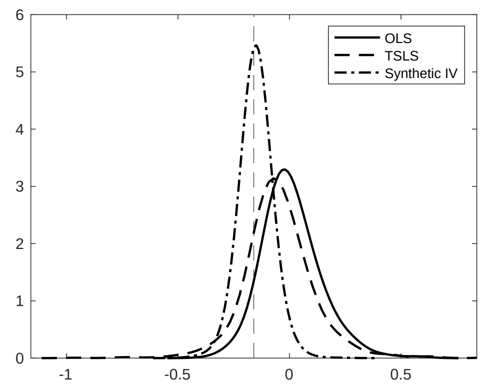
Figure 1.5: Model comparison in simulations



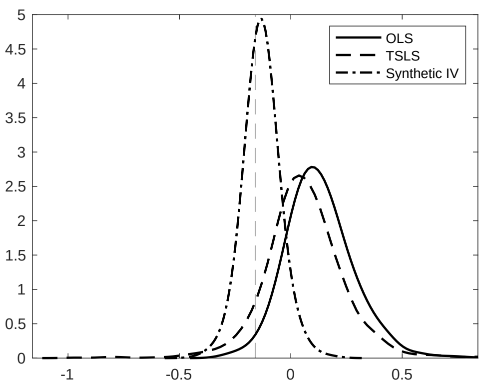
(a) $\rho = \rho_z = \rho_g = 0$



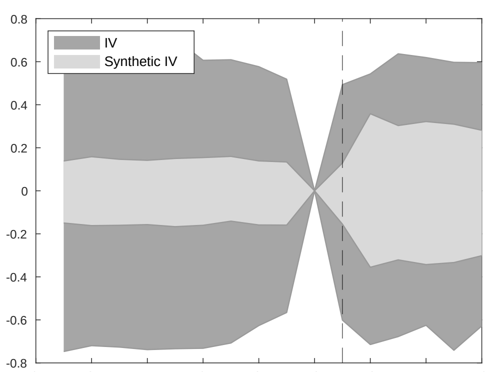
(b) $\rho = 0.5$, $\rho_z = \rho_g = 0$



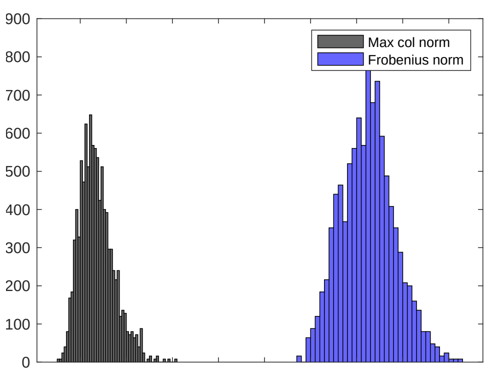
(c) $\rho = \rho_z = \rho_g = 0.5$



(d) $\rho = \rho_z = \rho_g = 0.7$



(e) Event study estimates



(f) Weight density

Note: Panels (a)-(d) display kernel density plots for TWFE OLS, TWFE TSLS and the synthetic IV. Panel (e) shows simulated event study estimates with 95% confidence bands for $\rho = \rho_z = \rho_g = 0.5$. Panel (f) shows a histogram of $\max_i \|w^i\|_1$ and $\sum_i \|w_i\|_2^2$ for $\rho = \rho_z = \rho_g = 0.5$. Simulations are done over 10000 iterations with the parameters calibrated to the Syrian example.

on the other hand is approximately unbiased, and we can reject that the estimated effect is zero at the 5% significance level. When we increase the correlation to $\rho = \rho_z = \rho_g = 0.7$ in panel (d), the bias in the OLS and TSLS estimators increases, but the synthetic IV continues to exhibit close to zero bias.

To relate this simulation results to the 'pre-trends' discussion in Figure 1.5, panel (e) shows the event study coefficients (over 10000 simulations). Before the treatment starts at $T_0 = 10$, the coefficients should be close to zero, as the instrument is not active. After the treatment starts the variation in the event study coefficients should increase given that $|\theta| > 0$.[8] The TSLS estimator shows large deviations in the pre period similar to those in the post period, indicating the presence of 'pre-trends' and unmeasured confounding. On the other hand, the synthetic IV limits the deviations in the pre period and also reduces the post period variation, suggesting that it is partialling out part of the unmeasured confounding. Finally, panel (f) in Figure 1.5 shows the density of the estimated synthetic control weights when $\rho = \rho_z = \rho_g = 0.5$. The max col norm histogram shows the value of $\max_i \|w^i\|_1 = \sum_{j \neq i} |w_{ji}|$ across simulations. It shows that the weights are dense, with no one unit receiving all weight across synthetic controls, and that the weight condition in Theorem 4 is likely to be satisfied as $\frac{1}{\sqrt{JT_1}} \max_i \sum_{j \neq i} |w_{ji}| \simeq 2.5/\sqrt{26 \times 6} \simeq 0.2$.

To gain further insight into the behavior of the different estimators, Table 1.1 shows the mean, variance, bias and mean-squared error of each estimator over 10000 simulations for different correlations and noise levels. In all cases, the SIV outperforms the OLS and TSLS (with two-way fixed effects) in terms of bias and mean-squared error, often by an order of magnitude. Furthermore, the SIV exhibits close to zero bias in settings with moderate noise and correlation levels. We compare the SIV to the synthetic IV for which we only debias the instrument $\tilde{\theta}_Z$ (SIV Z in the table), the projected SIV and the ensemble estimator proposed in section 1.5. As expected from the theoretical discussion, while $\tilde{\theta}_Z$ performs similar to SIV for moderate correlation settings, as the correlation grows (Q becomes smaller) the finite sample behavior of the SIV is better and $\tilde{\theta}_Z$ exhibits more bias. More so, also as expected, the projected SIV performs worse than the SIV in low noise settings and high correlation levels, but is robust to increasing the noise level. Intuitively, the projected SIV estimates (and TSLS estimates) do not change much as σ_ϵ is increased as the noise is orthogonal to the instrument. With this in mind, the ensemble estimator that combines the SIV and the projected SIV achieves better bias and MSE than the SIV in most settings.

In section A.1.10 in the appendix we compare the SIV estimator with an estimator that

[8] Note that in this design across simulations we expect that the event study coefficients are centered at zero, which is why both IV and SIV are centered at zero. However, within a simulation the coefficients should be zero in the pre-period and non-zero in the post-period. Hence, the variation across simulations gives speaks to the bias in the coefficients in the pre and post periods.

combines the iterative procedure of Bai (2009) and PCA to directly estimate and project out the factor structure. We find that directly estimating the factor structure leads to MSEs comparable to that of the SIV estimator in cases with a small number of factors (with the SIV exhibiting smaller bias). However, as the number of factors to estimate increases, the SIV estimator performance only deteriorates slightly, while the PCA based procedure becomes increasingly biased towards the TSLS estimator. This finding mirrors the discussion and simulations of Imbens and Viviano (2023) in which synthetic control procedures are shown to offer finite sample performance improvements relative to direct factor model estimation.

Table 1.1: Simulations calibrated to the Syrian example for different $\rho = \rho_z = \rho_g = r$ and σ_ϵ.

	r=0.5				r=0.7				r=0.9			
	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE
					$\sigma_\epsilon = 1/2\sigma_{Syria}$							
OLS (TWFE)	-0.018	0.018	0.142	0.038	0.096	0.024	0.256	0.090	0.244	0.023	0.404	0.186
TSLS (TWFE)	-0.049	0.023	0.111	0.035	0.058	0.031	0.218	0.078	0.201	0.028	0.361	0.158
SIV	-0.155	0.002	0.005	0.002	-0.142	0.003	0.018	0.003	-0.087	0.008	0.073	0.013
projected SIV	-0.187	0.009	-0.027	0.010	-0.173	0.008	-0.013	0.008	-0.113	0.012	0.047	0.014
SIV + projected	-0.156	0.002	0.004	0.002	-0.144	0.003	0.016	0.003	-0.092	0.008	0.068	0.012
SIV Z	-0.142	0.003	0.018	0.004	-0.108	0.006	0.052	0.008	0.003	0.010	0.163	0.037
					$\sigma_\epsilon = \sigma_{Syria}$							
OLS (TWFE)	0.005	0.017	0.165	0.044	0.126	0.022	0.286	0.104	0.277	0.021	0.437	0.212
TSLS (TWFE)	-0.049	0.023	0.111	0.036	0.058	0.032	0.218	0.079	0.200	0.028	0.360	0.158
SIV	-0.151	0.004	0.009	0.004	-0.132	0.005	0.028	0.006	-0.056	0.013	0.104	0.024
projected SIV	-0.188	0.011	-0.028	0.012	-0.169	0.011	-0.009	0.011	-0.092	0.018	0.068	0.023
SIV + projected	-0.154	0.004	0.006	0.004	-0.137	0.005	0.023	0.006	-0.067	0.013	0.093	0.021
SIV Z	-0.135	0.005	0.025	0.005	-0.092	0.007	0.068	0.012	0.037	0.011	0.197	0.050
					$\sigma_\epsilon = 2\sigma_{Syria}$							
OLS (TWFE)	0.041	0.015	0.201	0.056	0.170	0.021	0.330	0.130	0.327	0.020	0.487	0.257
TSLS (TWFE)	-0.049	0.025	0.111	0.038	0.057	0.033	0.217	0.080	0.199	0.030	0.359	0.159
SIV	-0.144	0.008	0.016	0.008	-0.114	0.010	0.046	0.012	-0.013	0.020	0.147	0.042
projected SIV	-0.189	0.017	-0.029	0.018	-0.163	0.018	-0.003	0.018	-0.064	0.035	0.096	0.044
SIV + projected	-0.151	0.007	0.009	0.007	-0.125	0.009	0.035	0.011	-0.031	0.022	0.129	0.039
SIV Z	-0.125	0.008	0.035	0.009	-0.070	0.011	0.090	0.019	0.075	0.013	0.235	0.068
					$\sigma_\epsilon = 4\sigma_{Syria}$							
OLS (TWFE)	0.090	0.014	0.250	0.077	0.231	0.019	0.391	0.171	0.395	0.019	0.555	0.327
TSLS (TWFE)	-0.050	0.029	0.110	0.041	0.056	0.037	0.216	0.083	0.198	0.032	0.358	0.160
SIV	-0.132	0.027	0.028	0.028	-0.090	0.019	0.070	0.023	0.037	0.030	0.197	0.069
projected SIV	-0.187	0.029	-0.027	0.029	-0.152	0.034	0.008	0.034	-0.047	0.566	0.113	0.579
SIV + projected	-0.144	0.022	0.016	0.022	-0.109	0.019	0.051	0.022	0.000	0.200	0.160	0.226
SIV Z	-0.112	0.014	0.048	0.017	-0.045	0.017	0.115	0.030	0.110	0.017	0.270	0.090
					$\sigma_\epsilon = 8\sigma_{Syria}$							
OLS (TWFE)	0.147	0.013	0.307	0.107	0.302	0.016	0.462	0.229	0.474	0.016	0.634	0.418
TSLS (TWFE)	-0.053	0.038	0.107	0.050	0.052	0.047	0.212	0.092	0.193	0.042	0.353	0.167
SIV	-0.124	0.031	0.036	0.032	-0.062	0.041	0.098	0.050	0.081	0.049	0.241	0.107
projected SIV	-0.181	0.060	-0.021	0.061	-0.155	0.513	0.005	0.513	0.036	2.163	0.196	2.201
SIV + projected	-0.141	0.029	0.019	0.030	-0.095	0.093	0.065	0.097	0.063	0.486	0.223	0.536
SIV Z	-0.100	0.025	0.060	0.028	-0.020	0.028	0.140	0.047	0.139	0.024	0.299	0.113

Table 1.2 considers inference for the SIV estimator in our calibrated simulation design. The upper panel checks the coverage under the null $\theta = 0$ in the setting in which the common factors are exchangeable ($\kappa = 0$). We average over 10000 simulations the coverage of the 95% confidence intervals constructed according to 1.10 for the SIV estimator using the estimated

variance $\tilde{\sigma}^2_{TSLS}$, the permutation p-value ($\hat{p}$) detailed in section 1.5 (with $T_b = 8$) and the weight density $\max_i \|w^i\|_1$ from the condition in Theorem 4. Overall, despite the small number of time periods and units ($T_1 = 6$, $T_0 = 10$, $J = 6$), CI based on the variance estimator $\tilde{\sigma}^2_{TSLS}$ exhibits close to nominal size in settings with moderate noise and moderate correlation with the unobserved confounder. In cases with low noise level and low correlation, the SIV exhibits slight over-coverage (as $\tilde{\sigma}^2_{TSLS}$ is inflated by the $\tilde{\alpha}_{it}$ terms), and in cases with high noise level and correlation, under-coverage. The under-coverage and finite sample bias for high correlation cases is both due to the rate of convergence depending on the correlation and a weaker first stage due to removing variation from the instrument (smaller Q). The permutation p-value, by construction, has nominal size regardless of the correlation or noise level, as can be seen by the average $\hat{p}$ being 0.5 across simulation designs. On the other hand, once we check the power of the test to detect the small $\theta = -0.16$ effect in the lower panel of Table 1.2, we find that the standard CI exhibit good power (60-90%) except in high noise cases, while the permutation p-value is under-powered. This should not be surprising given the small number of time periods and small effect. Overall, the good performance of the standard CI for the SIV estimator, in contrast to the theoretical rates, can be attributed to the synthetic controls balancing as the noise level increases. This can be seen by observing that the weight condition (given by $\max_i \|w^i\|_1$) improves with σ_ϵ and remains bounded for all simulation designs.

Table 1.2: Size and power simulations for $\rho = \rho_z = \rho_g = r$ and $\sigma_\epsilon = \iota\sigma_{Syria}$.

				$H_0 : \theta = 0$ ($\kappa = 0$)					
		r=0.0			r=0.5			r=0.7	
	Coverage	$\max_i \|w^i\|_1$	$\hat{p}$	Coverage	$\max_i \|w^i\|_1$	$\hat{p}$	Coverage	$\max_i \|w^i\|_1$	$\hat{p}$
ι									
0.100	0.998	2.585	0.508	0.997	2.595	0.489	0.990	2.570	0.495
0.500	0.986	2.418	0.510	0.963	2.419	0.493	0.950	2.405	0.498
1.000	0.956	2.322	0.509	0.938	2.318	0.496	0.898	2.305	0.499
2.000	0.912	2.213	0.499	0.890	2.210	0.498	0.822	2.202	0.502
4.000	0.878	2.123	0.504	0.829	2.119	0.495	0.732	2.117	0.511
8.00	0.847	2.061	0.508	0.800	2.059	0.499	0.675	2.060	0.501

				$H_1 : \theta = -0.16$ ($\kappa = 0.5$)					
		r=0.0			r=0.5			r=0.7	
	Power	$\max_i \|w^i\|_1$	$\hat{p} < 0.05$	Power	$\max_i \|w^i\|_1$	$\hat{p} < 0.05$	Power	$\max_i \|w^i\|_1$	$\hat{p} < 0.05$
ι									
0.100	0.910	2.593	0.568	0.865	2.609	0.502	0.855	2.602	0.410
0.500	0.865	2.439	0.294	0.794	2.442	0.260	0.788	2.455	0.187
1.000	0.813	2.349	0.203	0.707	2.347	0.178	0.712	2.360	0.130
2.000	0.726	2.247	0.150	0.592	2.247	0.127	0.592	2.252	0.104
4.000	0.619	2.153	0.114	0.447	2.150	0.090	0.455	2.154	0.086
8.00	0.490	2.085	0.088	0.347	2.081	0.090	0.319	2.080	0.088

The key takeaways from the simulations are that the SIV performs well in cases in which the TSLS and OLS do not, but that the estimator may be biased when the signal-to-noise level is weak (high noise) or the correlation with the unmeasured confounder is very large (no first stage). Another aspect highlighted by the theory and by reviews of best practices for synthetic control estimators (Abadie and Vives-i-Bastida, 2025), is how the relative sizes of T_0, T_1 and J influence the behavior of the estimator. The consistency result requires requires that both JT_1 is large and $\sqrt{J/T_0}$ is small. In our baseline simulation we considered a setting with $JT_1 = 156$ and $\sqrt{J/T_0} = 2.6$, but with a strong instrument and signal-to-noise ratio. In the appendix, we consider alternative simulation designs with different number of time periods, units, and weaker instrument and signal-to-noise ratios. Overall, for the different designs the same conclusions are drawn and the SIV estimator consistently outperforms the TSLS and OLS estimators (with two-way fixed effects).

With the simulation results in mind, we propose four robustness checks that practitioners should implement when using synthetic IV or similar estimators:

1. **Checking your first stage**: the debiasing procedure leads to a weaker first stage as variation is removed from the instrument. In cases with strong correlation between the instrument and the confounder (small Q) if the synthetic IV estimator exhibits a weak first stage researchers should be worried about using an IV strategy and the synthetic estimator.

2. **Checking your pre-treatment fit**: if the debiased outcomes exhibits large deviations in the pre-treatment period or an event study design reveals pre-trends, it is likely that the synthetic estimator will be biased and the signal-to-noise level too large for the estimator to perform well in the researcher's sample.

3. **Back testing**: given that the finite sample bias depends on the expected pre-treatment fit, back testing the intervention an evaluating the fit of the estimator, with weights computed in a training period, on a blank period can reveal whether the good pre-treatment fit was due to over-fitting (high noise) or due to partialling out the confounder. Additionally, researchers may implement the proposed permutation p-value to test the sharp null that the pre and post treatment event study estimates come from the same distributions.

4. **Checking the weight density**: ensuring that the synthetic control weights are not disproportionately weighting a few units by looking at the distribution of weights can reveal whether the asymptotic normality approximation is likely to be good in the researcher's empirical setting.

In the following section we implement these robustness checks when re-evaluating the effect of the Syrian refugee crisis using the synthetic IV.

1.7 Empirical applications

1.7.1 Revisiting the Syrian refugee shock

With the SIV tool at our disposal, we now re-visit our analysis of the impact of Syrian refugees on the salaried employment of low-skill natives. As detailed in section 1.3 we first solve the synthetic control problem (1.3) using the demeaned data between 2004–2010.[9] Then, we create synthetic regions with outcome Y^{SC}, treatment R^{SC}, and instrument Z^{SC} and debias the data by subtracting the raw data with the synthetic data, generating $\tilde{Y}_{it}$, $\tilde{R}_{it}$ and $\tilde{Z}_{it}$.

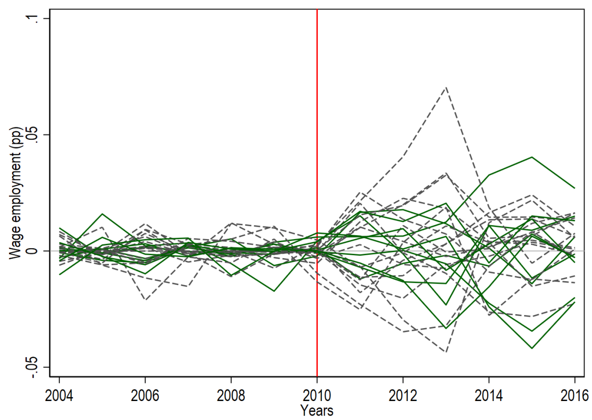
Before estimating the treatment effect via TSLS on the debiased data following step 2 in section 1.3, we implement the quality checks detailed in 1.6 to ensure that we are in a setting in which we can apply the SIV estimator. First, we check the matching quality in the pre-period, since as discussed in the theory section, goodness of fit is necessary to get consistent estimates using SIV. We plot the debiased wage-employment data ($\tilde{Y}$) in panel (a) Figure 1.6a, where black dashed lines belong to the less intensely treated regions that received less than 2% of refugees compared to their native population by 2016, and the green straight lines belong to the more intensely treated regions. During the training period 2004–2010, the debiased data is close to zero, which implies that we were able to match well on the pre-treatment trends.

The second check we perform is to look at the first-stage using the debiased data, as the partialling out procedure will make the first stage weaker. We plot the first-stage estimates in panel (b) of Figure 1.6b. In our case, the debiased data maintains a strong first-stage. In a regression of $\tilde{R}$ on $\tilde{Z}$ while controlling for two-way fixed effects, the F-stat is 218. The third check we perform is to look at the reduced-form using the debiased data. If the matching was successful, i.e., the donor pool had regions with similar trends for all the regions in the sample, then the event-study design on the debiased data should find estimates around zero in the pre-period. As in the discussion of Figure 1.4, we estimate event-study regressions as in (1.9) for Z_i and $\tilde{Z}_i$ and plot the estimates Figure 1.6c. Adjusting for pre-rends, SIV finds slightly stronger disemployment effects in the post-period.

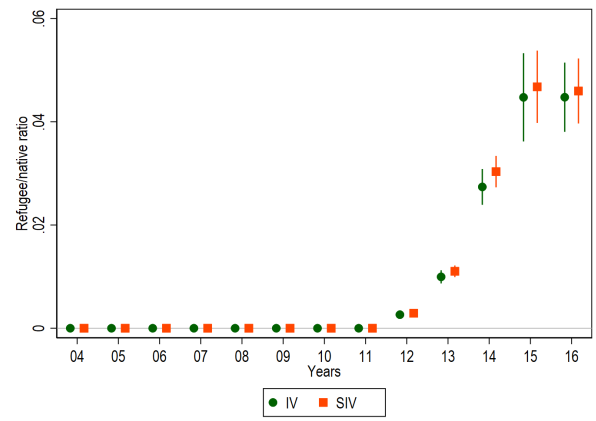
To test for over-fitting bias, we perform back-testing. In particular, instead of using the

[9] Demeaning the individual regions is an important detail in the Turkish setting due to the large heterogeneity in development rates across regions. For example, Istanbul is the most developed region with the highest employment rate in Turkey. No convex combination of other regions can match Istanbul on levels, but matching on trends is feasible.

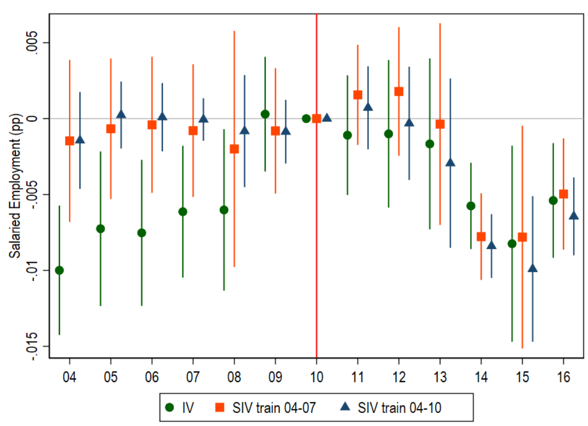
Figure 1.6: Quality Checks



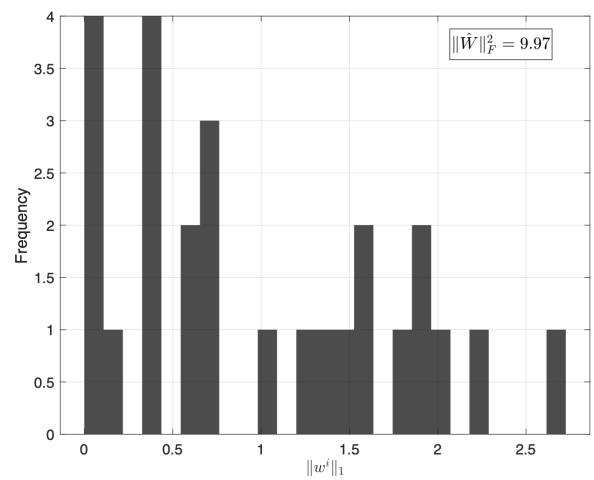
(a) Debiased Data



(b) Debiased First-Stage



(c) Debiased Reduced-Form



(d) Weight density

Notes: Panel (a) uses the debiased data. The green solid lines belong to the intensely treated regions, the black dashed lines belong to the rest, and the cutoff is 2% refugee/native ratio. The first-stage using both raw and debiased data is plotted in Panel (b). The F-stat in the main first-stage is 154 with the raw data and 218 with the debiased data. In Panel (c), the reduced-form estimates come from the event-study design shown in equation (1.9). The outcome variable is the wage-employment rate of low-skill natives. Panel (d) shows $\|w^i\|_1$ for each Turkish region i, where w^i are the SC weights assigned to region i in the SC of the other regions. Standard errors are clustered at the region level. The 95% confidence interval is plotted.

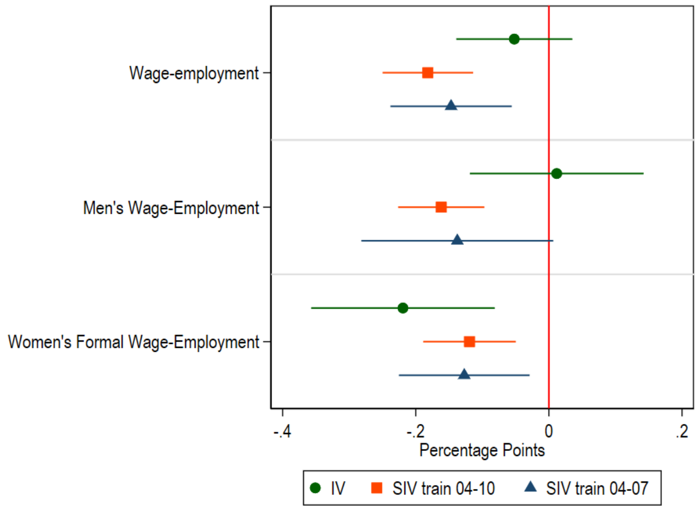
entire pre-period in the matching, we solve for the SC weights using data between 2004-2007 and follow the rest of the algorithm as specified before. We plot the estimates in Figure 1.6c in blue. Despite the reduced amount of time periods that we match on, the reduced form does not find any placebo effect in the pre-period. All the estimates between 2004–2010 are both quantitatively close to zero and statistically not significant, meaning synthetic distance is successfully capturing the unmeasured confounder in the pre-period. Furthermore, we compute the permutation p-value described in section 1.5 and find that $\hat{p} = 0.023$, meaning that we can reject at the 5% significance level that the post-treatment event study coefficients are zero and jointly equivalent to the pre-treatment coefficients.

Finally, Figure 1.6d shows the l_1 norm of the weights each region has across synthetic controls. Given that no one region receives a large amount of the weight, and the maximum value (2.6) is small relative to $\sqrt{JT_1} \approx 12.5$, we are confident that the density condition of Theorem 4 is satisfied.

It is worth further discussing why IV and SIV estimates differ less in the post-period than in the pre-period in Figure 1.4 and Figure 1.6c. While it is impossible to know the exact nature of the unmeasured confounding, some likely explanations can help understand the nature of the pre-trend. As explained in Gulek (2025), the regions close to the Syrian border are less-developed than the rest. Between 2004–2010, Turkey's GDP per capita grew by 75%. The data suggests that the less developed southeast regions were "catching up" to the rest of Turkey with higher salaried employment growth rates. This aggregate growth period did not last as Turkey entered a recession in 2013. If economic growth in the pre-period was the main reason behind the pre-trends, it is likely that these pre-trends would not extrapolate into the post-period. The SIV is capturing this underlying change in the unobserved confounder by not changing the post estimates by a considerable margin. The presence of pre-trends does not necessarily imply a violation of the parallel trends assumption in the post-period. Although this principle is widely recognized in theoretical discourse, it often receives insufficient attention in empirical studies. This oversight may stem from the challenge of publishing research that identifies significant pre-trends without addressing them through parametric techniques, like controlling for linear trends. In our case, adjusting for linear trends would have caused us to overestimate immigrants' effect on natives' salaried jobs.

The degree by which SIV and IV estimates differ in the post depends on the persistence of the unobserved confounders. For different outcomes, IV and SIV estimates can differ more. To show this in our context, we estimate event studies for the immigrants' effect on the salaried employment of low-skill men and formal salaried employment of low-skill women. Figure A.3, in the appendix, plots the estimates. In the pre-period, whereas regions close to

Figure 1.7: SIV and IV estimates.



Notes: This Figure plots the IV estimates ($\hat{\theta}_{TSLS}$) in green circles, the SIV estimates ($\tilde{\theta}$) in orange squares, and the backdated SIV estimates in blue triangles for the three main outcome measures. 95% confidence intervals are provided, in the case of SIV they are constructed using $\tilde{\sigma}_{TSLS}$. The permutation p-values ($\hat{p}$) are 0.023, 0.013 and 0.032 respectively.

border are observing a relative increase in men's salaried employment as seen in panel (a), they observe a relative decrease in women's formal salaried employment as seen in panel (b). In both cases, SIV eliminates the pre-trends and adjusts the estimates in the post period.

Having seen how SIV addresses the pre-trend problem in the event-study designs and satisfies the recommended empirical checks, we continue by implementing the second step of the algorithm: we apply TSLS to the debiased data. We estimate the effect of Syrian refugees on low-skill natives' wage-employment, and show heterogeneity by sex and formality. We plot the estimates in Figure 1.7. As a benchmark, we first show the IV estimates. A researcher using IV would find no effect on natives' or men's salaried employment, and find negative effects on women's salaried employment. However, using SIV, we find that Syrian refugees lowered natives' salaried employment in all cases. A 1 pp increase in the refugee/native ratio decreases low-skill natives' salaried employment rate by 0.16 pp for men and 0.10 pp for women. As a robustness check, we also show the results that rely on estimated weights using the only 2004–2007 as a training period. The results remain quantitatively and qualitatively very similar. Accordingly, the permutation p-values reject the null effect ($\theta = 0$) for all three outcome measures at the 5% significance level.

It is worth highlighting how much our method impacts the economic conclusions in the empirical setting. Turkey hosts the largest number of refugees in the world. Turkey's three most treated exposed regions (the ones that received the most refugees) observed an increase in labor supply of more than 10% in just five years. Refugees, especially men, have a high propensity to work: 87% of prime-age men are "employed" in Turkey (Turkish Red Crescent and WFP, 2019). Despite this *large* labor supply shock, in a short enough time period where spatial markets are unlikely to equilibrate and despite male refugees' having higher employment rates than male natives, the standard IV finds no disemployment effects for native men. Theoretically justifying this result would require either completely flat labor demand curves (Borjas, 2003) or refugees to provide a substantial positive product demand shock (Borjas, 2014). There is very little empirical evidence for both, especially considering that Syrian refugees left most of their wealth behind while escaping a civil war. SIV reveals that this significant labor supply shock has caused native disemployment in the short run for both men and women, which is consistent with economic theory.

1.7.2 Revisiting the China shock

SIV can be applied to any exposure and shift-share design. As an additional empirical example, we estimate the effect of Chinese imports on manufacturing employment in the United States following the identification strategy of Autor et al. (2013). The authors are interested in the following regression (where we omit covariates for simplicity)

$$\begin{aligned} Y_{it} &= \beta X_{it} + \epsilon_{it} \\ X_{it} &= \gamma Z_{it} + \eta_{it} \end{aligned} \tag{1.14}$$

where Y_{it} is the percentage point change in the manufacturing employment rate for region i in decade t, $X_{it} = \sum_k s_{ikt} g_{kt}^{US}$ is the import exposure, where s_{ikt} is the industry-location share at the beginning of period, and g_{kt} is a normalized measure of the growth of imports from China to the US in industry k. The import exposure to China is instrumented by the increase in Chinese imports by high-income countries: $Z_{it} = \sum_k s_{ikt-1} g_{kt}^{\text{high-income}}$, where s_{ikt-1} is the share of industry k in the previous period and $g_{kt}^{\text{high-income}}$ is a normalized measure of the growth of Chinese imports to selected high-income countries. We focus on the TSLS estimates from Tables 2 and 3 of Autor et al. (2013).[10]

The paper considers 4 periods of data: 1970–1980, 1980–1990, 1990–2000, 2000-2007. We denote these periods by their starting year throughout the exercise (e.g., 1990 refers to the period between 1990–2000). The Chinese import shock takes place in 1990–2000 and 2000–2007. The growth in Chinese imports from high-income countries in 1990–2000, $g_{k,1990}^{\text{high-income}}$, predicts an exposure across US commuting zones via their pre-existing industry

[10] In Table 2, the IV estimates without covariates and some placebo checks are shown. Table 3 shows the regressions with additional covariates.

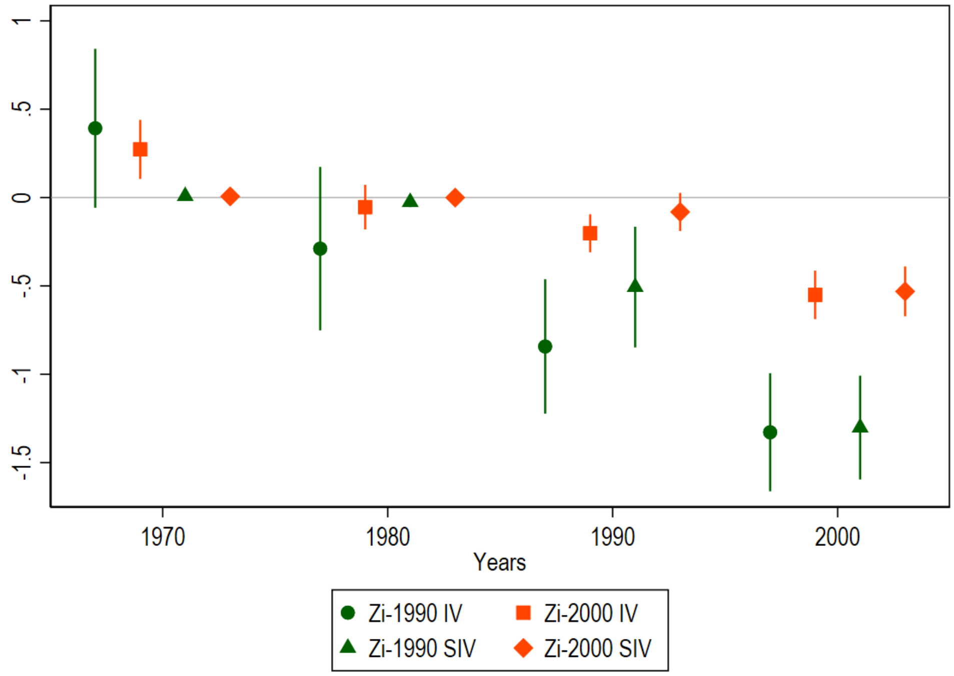
structure. We denote this exposure as $Z_{i,1990}$ and define $Z_{i,2000}$ similarly, as the exposure predicted by the growth in Chinese imports from high-income countries between 2000–2007. Together, $Z_{i,1990}$ and $Z_{i,2000}$ constitute the shift-share instrument used as $Z_{it} = Z_{i,1990}\mathbb{1}(t = 1990) + Z_{i,2000}\mathbb{1}(t = 2000)$. These two exposure measures have a correlation of 0.67 across 722 commuting zones, which implies that regions that have a high exposure in 1990 were also likely to have an high exposure in 2000. This suggests that the trade shocks in 1990 and 2000 were unlikely to be i.i.d., and hence we follow the exogeneity of shares assumption in this shift-share design (Goldsmith-Pinkham et al., 2020) as opposed to the exogeneity of shifts assumption in Borusyak et al. (2022).

Dissecting the shift-share instrument Z_{it} into its two "exposure" components $Z_{i,1990}$ and $Z_{i,2000}$ lends itself to an event-study design, as we already considered in section 1.2 (equation (1.2). Figure 1.3 shows that the exposure shares predict decreases in manufacturing employment in both 1990 and 2000, which is one of the core results in the China shock paper. However, this figure also reveals that the correlation between the exposure shares and manufacturing growth was positive in 1970, two decades before the Chinese shock, and has been decreasing since then. For example, the coefficient estimate of $Z_{i,1990}$ goes from 0.39 in 1970 to -0.28 in 1980. This pre-trend raises a concern regarding the validity of the *exogeneity of shares* assumption in this shift-share design because if this trend was to continue absent the China shock, we would have estimated the same "negative" employment effects in 1990 and 2000.

To apply the SIV estimator to the China shock example we follow the algorithm described in section 1.3. We first solve the synthetic control problem for all of the 722 CZs, where we match on the growth rate between 1970 and 1980, Then, we obtain the synthetic variables, y^{SC}, x^{SC}, Z^{SC}, and compute the debiased values $\tilde{Y}, \tilde{X}, \tilde{Z}$. Due to the small number of pre-periods and large number of donor units, the pre-treatment fit is almost perfect as can be seen in Figure 1.8 panel (b). As discussed in Abadie and Vives-i-Bastida (2025) in these settings it is likely that the synthetic control is fitting the noise, leading to over-fitting bias. To address this in the appendix we re-do the analysis limiting the donor pool to the closest 100, 50, 30, and 20 donor regions according to the Euclidean distance.

We investigate the effects of the China shock on US manufacturing by comparing the IV and the SIV estimates. We find that the SIV results are slightly smaller in magnitude, but overall similar to the IV findings. Looking at 1990 in Figure 1.8 panel (b), we see that the SIV finds a statistically significant decrease in manufacturing employment due to Chinese imports, but the coefficient estimates are slightly smaller in magnitude than the IV estimates. In 2000, on the other hand, we find quantitatively the same results as the original study: adjusting for the pre-trends in 1970 and 1980 does not meaningfully change the estimates in

Figure 1.8: Reduced-form estimates using the 1990 and 2000 shares



2000-2007.

Table 1.3: China shock effect

	1990–2000	2000-2007	1990–2007
	(1)	(2)	(3)
2SLS	-0.888	-0.718	-0.746
	(0.181)	(0.064)	(0.068)
SIV	-0.588	-0.726	-0.703
	(0.198)	(0.070)	(0.067)

Notes: The first row replicates columns 1–3 of Table 2 in Autor et al. (2013). The second row presents the estimates using the SIV. The SC weights are estimated using the manufacturing growth rates in 1970 and 1980.

Table 1.3 replicates the main findings in Autor et al. (2013). For the 1990–2000 period, the TSLS estimates suggest that a $1,000 increase in import exposure per worker leads to a decline in manufacturing employment of 0.89 pp. The SIV estimate for the same period is 33% smaller, but not statistically different, at 0.59 pp. This confirms the intuition from Figure

1.8 that adjusting for the pre-trend reduces slightly the China shock effect in 1990-2000. The results for the 2000-2007 period and the two periods combined (1990-2007) imply a decrease between 0.7 and 0.75 pp with little differences between TSLS and SIV. Autor et al. (2013) also report IV estimates with additional covariates. We replicate their results in the appendix in Table A.6 and find very comparable results between IV and SIV, revealing that the unmeasured confounding might be well proxied by the set of controls.

Overall, our replication implies that despite the strong pre-trend between 1970–1990, the IV estimates in 1990–2007 may not suffer from large biases due to unmeasured confounding. A potential explanation for this is that the decline in manufacturing employment growth suggested by the pre-trends flattens over time and disappears in the 2000s, leading to the small adjustment in the 1990-2000 period and no adjustment in the 2000-2007 period as reported by our SIV estimates.

1.7.3 The effect of search rankings

In this section we study an important question in platform and digital economics; what is the effect of product rankings on producer outcomes? Many digital platforms guide consumers to producers through search walls in which products are ranked according to ordered lists generated by a recommendation system. Some examples include search engines (e.g. Google, Bing), product market places (e.g. Amazon, Wayfair, Alibaba), food delivery platforms (e.g. Deliveroo, Uber Eats, Door Dash), travel comparison sites (e.g. Expedia, Google flights, Kayak) or social medial platforms (e.g. Facebook, Instagram, Tiktok). Given the importance of digital platforms, there is a growing literature in economics on the impact that rankings have on producer and consumer choices (e.g., Athey and Ellison (2011), Ursu (2018), Choi and Mela (2019), Hodgson and Lewis (2023), Compiani et al. (2024), Reimers and Waldfogel (2023)).

A key empirical challenge in the literature is that of estimating the causal effect of changing the rank of a producer from observational data. Given that ranking and recommendation systems are based on producer and consumer characteristics, observed ranks are endogenous. For example, top search results in Google search are determined through an auction, and therefore, advertisers with higher willingness to pay (which may be higher quality) are placed higher. Similarly, in food delivery platforms and social media platforms recommendation engines are trained on consumer and producer histories and therefore higher quality producers may receive higher rankings. This problem is well understood in the literature and there are a number of existing approaches to correct the endogeneity bias. For example, De los Santos and Koulayev (2017) consider a control function approach, Narayanan and Kalyanam

(2015) and Moehring (2024) consider a regression discontinuity design, Ghose et al. (2014) specify and estimate a simultaneous equation model, and Rutz et al. (2012) consider a latent instrumental variables method. However, the validity of the different methods often relies on strong assumptions, or specific empirical settings in which the recommendation assignment function is known, which is why Ursu (2018) opts to analyze the results of a ranking A/B test in which ranks are fully randomized. In this paper, we propose a new set of instruments that can be used to study this question, show how the synthetic IV estimator can be applied to deal with unmeasured confounding and validate our observational results using an A/B test.

We provide new estimates of the effect of rank on producer sales in the empirical setting of a food delivery platform that selectively offers preferential search contracts to some producers. Producers with preferential search contracts see their position in the overall search wall of the platform fixed near the top (e.g. fixed at position 2,3,5,7,11) for a specific period of time. We use this type of *fixed* contracts as an instrument for rank. Given that many platforms offer preferential deals, sponsored slots, or promotion campaigns to producers, we believe that these kind of IVs might be available in many important digital settings. The fixed contract IV may be good for several reasons. First, given that the contracts directly change the ranks, the instrument is likely very relevant. Second, given that the contracts, in general, do not change other aspects of the producer's offering[11] the exclusion restriction may hold. However, it is unlikely that the producers offered the contracts where chosen at random, therefore the instrument may be invalid due to unmeasured confounding. This situation is common in IV approaches to ranking systems. Rutz et al. (2012) note that common instruments used in these settings (e.g. lagged rankings or lagged outcomes) are unlikely to be valid due to omitted variable biases. In our setting, however, there is some hope. While the *fixed* contract IV may suffer from unmeasured confounding like the lagged IVs, we have at our disposal a pre-treatment period for each producer; the period before the producers were given a contract. Therefore, we are in a setting in which we may apply our proposed synthetic IV estimator.

Our data, provided by a large global delivery platform, consists of a sample of one million food delivery orders across six European cities between December 2022 and March 2023, involving 240 thousand customers and 1633 stores of which 183 received a fixed position contract (about 11%). The platform operates a marketplace for food delivery in which restaurants/stores are ranked in a search wall that customers browse before placing an order. The order/rank of the producers in the search wall depends on a recommendation system (unknown to the econometrician) involving producer specific characteristics and histories as well as order specific variables (such as geographical distance). Customers must scroll down

[11] In some cases commission fees are also simultaneously changed with the contract, but limited price pass-through is observed. See Vives-i-Bastida and Sabal (2024) for a model of the platform.

to discover new producers, with only a few producers being shown at any given time on the screen (4 or 5 depending on the device used). The sample is chosen to include an A/B test the platform performed in which ranks were randomized at the customer level. In January and February 2023, 10936 customers where assigned to a treatment arm in which each week the rankings on their search wall were randomly generated within their market (city). We separate the A/B test data from the main dataset as a test set for validation. This allows us to compare different observational methods against a randomized experiment benchmark as in LaLonde (1986).

To study the effect of ranks on store outcomes, we aggregate the data at the store and week level. The aggregation can be micro-founded under a consumer discrete choice model in which consumer utility depends on common parameters that can be identified from aggregate market shares (Berry, 1994). For an in depth structural model of the platform, producer and consumer choices see Vives-i-Bastida and Sabal (2024) who use related data to study the welfare consequences of preferential contracts in platforms. The observed data consists of an outcome Y_{it}, the number of orders a producer i receives in week t, treatments R_{it}^k, the share of orders in which the producer was ranked in the top $k = 5, 8, 10$, and an instrument Z_{it} which consists of the share of orders in which the producer was *preferenced*. We do not consider producers with $Y_{it} = 0$ for some t, as we are interested in the effect of rankings for non-fringe producers that at least receive some orders on any given week. The instrument satisfies the requirements imposed by our design in Assumption 1, as $Z_{it} = 0$ for $t \leq T_0^i$, where T_0^i is the start of the contract period, but Z_{it} may vary both in time for $t > T_0^i$ and across i as producers have different preference contracts that vary across time. Therefore, this is an example of an instrument Z_{it} that does not have a factor structure as in the shift-share designs considered for the Syrian refugee and China shock studies. It is also important to note that in this context it is likely that SUTVA is violated, as receiving a higher rank implies other stores receive a lower rank, another reason to aggregate outcomes at the store-week level is to mitigate this concern. We note however, that the estimates of all of our estimators will suffer from this issue, making comparisons across estimators valid, and that we see our estimated effects as lower bounds on the true causal effect.

As in the other empirical examples we can assess the validity of the fixed contract instrument Z_{it} by checking the first stage strength and by inspecting the event study based on the reduced form. The first stage regression of R_{it}^k on Z_{it} (with two-way fixed effects) shows that, as expected, the instrument is a strong predictor of the rank of a store (see panel (a) of Figure A.5 in the appendix). Having a fixed contract increases the share of orders in which the store ranks in the top 5 by about 8%, in the top 8 by about 22% and in the top 10 by about 46%, with the first stage F-statistics being respectively, 158, 260 and

287. Therefore, we are confident that the fixed contract instrument is strong. To assess the potential unmeasured confounding in the reduced form, we estimate the following event study regression

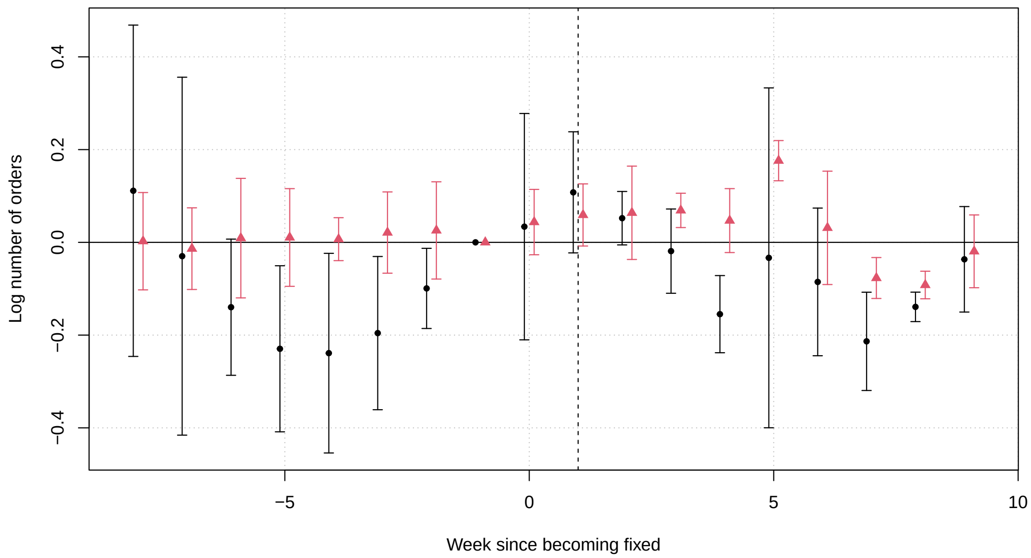
$$log(Y_{it}) = \alpha_i + \delta_t + \sum_k \theta_k \mathbf{1}(t - T_0^i = k) + \epsilon_{it},$$

where T_0^i denotes the time of the start of the fixed contract by producer i. This regression does not use the variation across units and time of the *intensity* of the preferential contract, but it highlights the potential unmeasured confounders in determining which producers are preferenced and which are not. Given that adoption of fixed contracts is staggered, we use the imputation estimator of Sun and Abraham (2021) to account for potential dynamic effect contamination. Panel (a) of Figure 1.9 shows the estimated reduced form effects for the fixed contract IV in black. It can be appreciated that while becoming fixed seems to initially increase the number of orders by about 11% one period after treatment, there are clear non-linear (U-shaped) pre-trends, with similar patterns appearing in the post-treatment period. Intuitively, this may be due to the platform giving fixed contracts to stores that are in different trajectories, for example stores that are expected to increase their sales and are in an upward trend relative to others.

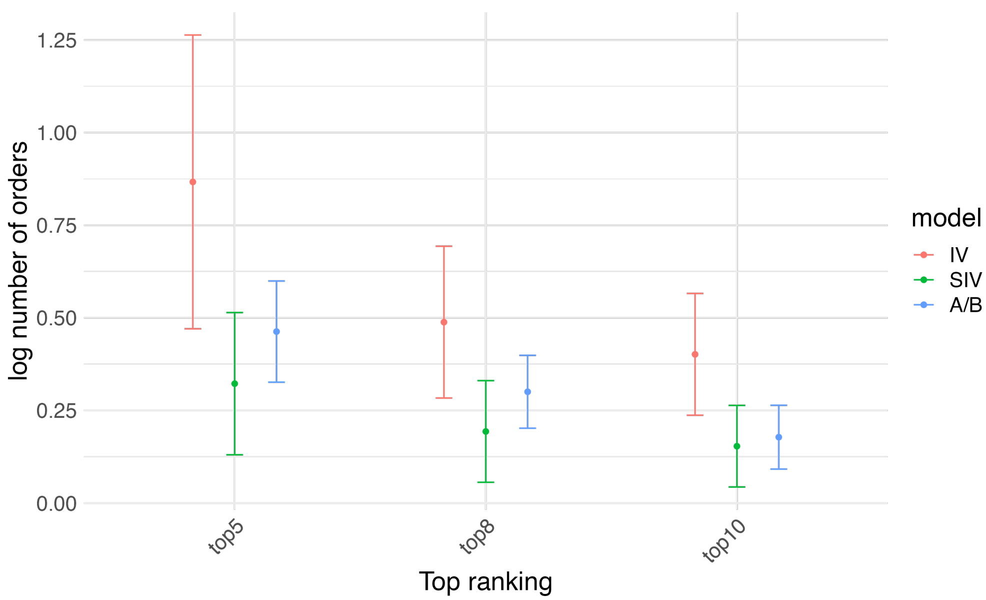
To deal for the unmeasured confounding highlighted by Figure 1.9 we apply the synthetic IV estimator. We modify the procedure described in section 1.3 to account for the staggered adoption of the fixed contracts. We compute the synthetic control weights for each fixed contract unit i using units that never receive a fixed contract and units that receive a fixed contract after unit i (i.e. units j such that $T_0^i \leq T_0^j$) as the donor pool. Then, we generate the debiased values $\tilde{Y}_{it}$, $\tilde{R}_{it}$ and $\tilde{Z}_{it}$ for $t > T_0^i$ and estimate the synthetic IV estimator for the post-treatment periods as in the standard case. Furthermore, given the large donor pool and number of treated units we both trim the donor pool to only match units close to each treated unit and drop the treated units with worst pre-treatment match. Our results, however, do not depend on this and are robust to different trimming operations. In panel (a) of Figure 1.9 we plot the synthetic IV event study estimates, using the debiased data, in red along with the standard IV estimates (in black). The synthetic IV does not exhibit pre-trends, as expected, but more importantly it also does not exhibit the U-shaped pattern in the post-treatment period that the standard IV does exhibit. The synthetic IV event study shows that after a producer becomes fixed the number of orders increases by about 10% over the following weeks.

Panel (b) of Figure 1.9 compares the fixed contract IV estimates, the synthetic IV estimates and the estimates from the A/B test. Given that in the A/B test, rank is randomized at the consumer level every week, the estimates should not suffer from omitted variable bias. We

Figure 1.9: Fixed contracts IV: reduced form and first stage



(a) Reduced form



(b) Rank effects estimates

Notes: Panel (a) shows the reduced form event studies for the fixed contract IV (black dots) and the synthetic IV (red triangles) using the Abraham and Sun (2020) estimator. The y-axis shows the number of orders in logs. Panel (b) shows the estimates of the effect of R_{it}^k for $k = 5, 8, 10$ on the log number of orders for the fixed contract IV, the synthetic IV and OLS for the A/B test sample. 95% confidence intervals are provided.

find that producer rank has a large effect on sales. The A/B test estimates (in blue) suggest that on average always being among the top 5 producers in the search wall on a given week leads to an increase in the number of orders of about 46% with respect to not being in the top 5. More so, as is common in the literature, we also find evidence of the effect decaying with the rank. Being in the top 8 leads to a smaller increase in sales of about 30%, and the effect for being in the top 10 is about 18%. When we compare the A/B test estimates to the fixed contract IV estimates, we see that there is likely a positive omitted variable bias despite the inclusion of two-way fixed effects. Respectively, the IV suggests a 86%, 49% and 40% increase in sales from being in the top 5, 8 and 10 ranks, more than double the A/B test estimates. This bias is in line with our reduced form discussion, since producers with higher growth potential may be more likely to receive a fixed contract. Reassuringly, the synthetic IV is successful in controlling the unmeasured confounding. The synthetic IV estimates (in green in Figure 1.9) are closer to the A/B test estimates, being respectively 33%, 18% and 14%. More importantly, a statistical test would not be able reject that the synthetic IV estimates are different from the A/B test for any of the rank treatments at a reasonable significance level. This is evidence that the proposed estimator appropriately controls the unmeasured confounding in a setting in which the IV estimator does not.

Overall, the application of the synthetic IV estimator to the rank effects example is relevant because (1) it shows how a new class of IVs can be used to study the question of how rankings affect producer outcomes in digital platforms (the fixed contract/promotion IVs) and (2) it shows that the synthetic IV estimator is effective in dealing with unmeasured confounding in a real setting in which we have access to an experiment to validate the observational estimates.

1.8 Conclusion

In this paper we provide a new method, the Synthetic IV, to deal with unmeasured confounding in panel data settings in which researchers have access to an instrumental variable that is only partially valid. By assuming a factor structure on the unobserved confounding term we derive conditions under which a synthetic IV estimator that combines Synthetic Controls and two-stage least squares is consistent and asymptotically normal. Through a simulation study, we show that the estimator performs well in a variety of empirical settings and removes the bias in cases in which TSLS and OLS with two-way fixed effects do not. We further showcase the applicability of SIV in three empirical examples: studying the effect of immigrants on labor markets using the Syrian refugee crisis in Turkey, studying the effect of Chinese imports on US manufacturing employment and studying the effects of rankings on producer outcomes in digital platforms.

Chapter 2

Bayesian and Frequentist Inference for Synthetic Controls

written jointly with Ignacio Martinez

2.1 Introduction

Synthetic control methods (Abadie and Gardeazabal (2003), Abadie et al. (2010)) are often used to estimate treatment effects of aggregate policy interventions. The method has been described as "arguably the most important innovation in the policy evaluation literature in the last 15 years" (Athey et al. (2021)). Despite this, statistical inference for synthetic controls is often non-trivial. To highlight the nature of the problem, consider the synthetic control problem. Suppose that we observe data for a unit that is affected by an intervention of interest and for a donor pool of untreated units. We are interested in estimating the treatment effect on the treated after the intervention. Synthetic controls match the treated unit to the weighted average of the donor units that most closely resembles the characteristics of the treated unit before the intervention. The treatment effect on the treated is then estimated by taking the differences in outcomes between the treated unit and the weighted average. Given that only one treated unit is available and that the synthetic control estimator depends on a weighted average of the donor units outcomes, it is not straightforward to provide asymptotic guarantees for inference on the treatment effect.

With this in mind, inference for synthetic controls has been approached from various angles. First, a body of literature has focused on permutation based inference relative to a benchmark assignment process, as pioneered in Abadie et al. (2010) and more recently in Firpo and Possebom (2018) and Abadie and L'Hour (2021) among others. Second, large sample properties of synthetic control estimators have been studied in linear factor model settings. Starting with Abadie et al. (2010) who provide a bound for the bias of synthetic controls and Ferman (2021) that derives asymptotic results under the assumption that the treated unit factor loading is in the convex hull of the donor units factor loadings as the number of pre-treatment periods and donor units grows. More recently, Imbens and Viviano (2023) and Arkhangelsky and Hirshberg (2023) consider the large sample properties of synthetic control estimators in settings with unmeasured confounding and selection into treatment. Other related research has focused on projection theory results for average treatment effects estimated by synthetic controls (Li (2020)) and panel data approaches that relax the simplex restriction (Hsiao et al. (2012)). Third, conformal inference procedures have also been proposed. For example, in Chernozhukov et al. (2021) under the assumption that the synthetic control estimator is as good as a true synthetic control that recovers the treated unit. Finally, Bayesian inference methods have also been studied. For instance, Pang et al. (2022) and Pinkney (2021) for linear factor model structures, Bayesian regression methods in Kim et al. (2020), Bayesian structural time series in Brodersen et al. (2015) and Scott and Varian (2014), or empirical Bayes approaches in Amjad et al. (2018).

A common thread in the literature is that either the synthetic control restriction that the weights are in the simplex is dropped, or the assumption is made that there exist a true synthetic control that is able to perfectly recover the treated unit. Motivated by this, in this paper we ask: *in linear factor models when is the minimizer of the statistical risk a synthetic control*? In a simple model, we derive conditions on the primitives of the factor structure (the factor loadings) that make the target parameter be in the simplex. Then, given this target parameter, we focus on frequentist and Bayesian inferential procedures. We show under which conditions the MLE can uniformly estimate the predictive part of the factor structure for the treated unit in post-treatment periods (without the error shock) and provide conditions for a Gaussian asymptotic approximation. Then, we propose a Bayesian synthetic control that preserves the main features of the standard model and derive a Bernstein-von Mises (BvM) style result to link the frequentist and Bayesian inference.

The BvM result may be of particular interest to applied researchers as it provides a new way to perform frequentist inference using synthetic controls that is computationally appealing. While the conditions for the result require a large number of pre-treatment periods relative to the number of donor units, we show through simulations that BvM convergence can be achieved for medium size panels even in sparse settings.

This paper contributes to the synthetic control literature in two ways. First, it provides a new characterization result for synthetic controls under linear factor models. Given that these models are often used to motivate synthetic controls, we see our result as relevant to applied researchers; we provide conditions under which factor models are well suited to study synthetic controls. Our main results reinforce the rule of thumb in the literature that without good pre-treatment fit synthetic control estimates may be misleading (Abadie (2021), Abadie and Vives-i-Bastida (2025)). However, we also expand the scope of the class of linear factor models that motivate synthetic controls. We show that the condition in Ferman (2021) that the treated unit factor loadings fall in the convex full of the donor unit factor loadings, while sufficient, might not be necessary; other factor structures may also motivate the use of synthetic controls. We see this a positive result, supporting design assumptions in synthetic control papers requiring the existence of weights that recover the true treated unit factor structure, as required for example in Chernozhukov et al. (2021) or as the limited confoundedness over units assumption enforces in Imbens and Viviano (2023).

Second, this paper contributes to the literature by providing a new inferential procedure. While other Bayesian synthetic control estimators have been proposed, implementations often drop the simplex assumption. We propose and justify theoretically a Bayesian synthetic control that preserves the simplex assumption. This feature is important for interpretability and to limit extrapolation. Importantly, it provides an easy way of evaluating if synthetic controls should be used to estimate a causal effect: by checking whether the synthetic treated unit can replicate the pre-treatment outcome of interest, the researcher is able to evaluate whether the synthetic control estimator is likely be biased, and whether their Bayesian model may be miss-specified. We implement our proposed Bayesian synthetic control in the *bsynth* R-package and provide additional features to help researchers understand the posterior treatement effect distribution and implicit weight estimates.

Finally, we apply the Bayesian procedure to a classic synthetic control application studied by Abadie et al. (2015), the German re-unification, and a more recent intervention; the attempted unilateral declaration of independence (UDI) of Catalonia in 2017. The Bayesian synthetic control yields a similar estimate to the frequentist one for the effect of the German re-unification on the GDP of West Germany. We find that the German re-unification lead to a 7.5% decrease of the GDP per capita. The effect of the UDI on Catalan GDP is smaller, but nevertheless, we find that the UDI lead to a 1% decrease in GDP. We posit, as Esteller-Moré and Rizzo (2022) note, that this decrease is due to capital reallocation caused by the increased political instability.

The paper proceeds as follows. Section 2 describes the standard frequentist synthetic control and conditions under which the statistical risk minimizer is a synthetic control and can be estimated consistently by MLE. Section 3 describes the Bayesian synthetic control, the Bayesian inference procedure, presents the Bernstein-von Mises result, the connection with frequentist inference through simulations and describes the *bsynth* R-package. Finally, Section 4 discusses the empirical application to the German re-unification and the Catalan secession movement.

2.2 The Frequentist Synthetic Control

2.2.1 Standard Synthetic Control for a single unit

Consider a setting in which we observe $J+1$ aggregate units for T periods. The outcome of interest is denoted by Y_{it} and only unit 1 is exposed to the intervention during periods $T_0+1,\ldots,T$. We are interested in estimating the treatment effect $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$ for $t > T_0$, where Y_{1t}^I and Y_{1t}^N denote the outcomes under the intervention and in absence of the intervention respectively. Since we do not observe Y_{1t}^N for $t > T_0$ we estimate τ_{1t} by building a counterfactual $\hat{Y}_{1t}^N$ of the treated unit's outcome in absence of the intervention.

As in the standard synthetic control our counterfactual outcome will be given by a weighted average of the donor units' outcomes, that is $\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}$ for a set of weights $\mathbf{w} = (w_2,\ldots,w_{J+1})'$. To choose the weight vector $\mathbf{w}$ we use observed characteristics of the units and pre-intervention measures of the outcome of interest. Formally, we let the $K \times 1$ design matrix for the treated unit be $\mathbf{X}_1 = (Z_1, \bar{Y}_1^{\mathbf{K}_1}, \ldots, \bar{Y}_1^{\mathbf{K}_M})'$, where $\{\bar{Y}_1^{\mathbf{K}_i}\}_1^M$ represent M linear combination of the outcome of interest for the pre-intervention period. Similarly, for the donor units, $\mathbf{X}_0$ is a $K \times J$ matrix constructed such that its jth column is given by $(Z_j, \bar{Y}_j^{\mathbf{K}_1}, \ldots, \bar{Y}_j^{\mathbf{K}_M})'$. We call the K rows of the design matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ the *predictors* of the outcome of interest. This can include, for example, lags of the outcome variable and important context dependent characteristics of the aggregate units averaged over the pre-treatment period.

Abadie et al. (2010) propose estimating the $\boldsymbol{w}$ by solving the following program:

$$\min_{\boldsymbol{w} \in \Delta^J} \|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{w}\|_V = \left(\sum_{h=1}^{k} v_h (X_{h1} - W_2 X_{h2} - \cdots - W_{J+1} X_{hJ+1})^2 \right)^{1/2},$$

where Δ^J denotes the $J-$dimensional simplex and the researcher can choose the predictor weighting matrix $V = \text{diag}(v_1, \ldots, v_k)$ using his domain knowledge or using a data-driven procedure to optimize pre-treatment fit.

Given our synthetic control $\hat{\mathbf{w}}$ we can estimate our treatment effect on the treated for $t > T_0$ by:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} \hat{w}_j Y_{jt}.$$

In the following section we motivate synthetic controls when the potential outcomes are given by linear factor models. We derive conditions under which the target parameter will be in the simplex for the case in which the design matrix includes only the pre-treatment periods and show how a MLE can estimate the treated factor structure as J and T_0 grow.

2.2.2 Linear factor model

We start by carefully examining identification and inference of the predictive part of the treated outcome in a simple factor model. Following Ferman (2021), Ferman and Pinto (2021) and Hsiao et al. (2012) we consider the following linear factor model for the potential outcomes:

$$Y_{it}(0) = \boldsymbol{\lambda}_i' \mathbf{F}_t + \epsilon_{it},$$
$$Y_{it}(1) = \tau_{it} + Y_{it}(0).$$

The observed data y_{it} is given by

$$y_{it} = d_{it}Y_{it}(1) + (1 - d_{it})Y_{it}(0),$$

and only the first unit is treated, so $d_{it} = 1$ for $i = 1$ and $t > T_0$ and $d_{it} = 0$ otherwise. We make the following simplifying assumptions:

(A1) – *factors*

(a) we have only one factor such that $\lambda_i, F_t \in \mathbb{R}$,

(b) $F_t \sim_{i.i.d} N(0, \sigma^2)$, where $0 < \sigma < \infty$,

(A2) – *idiosyncratic shocks*

(a) $\epsilon_{it} \sim_{i.i.d} N(0, \sigma_\epsilon^2)$, where $0 < \sigma_\epsilon < \infty$.

Under **A1-A2** it is shown in the appendix that the conditional distribution of Y_{1t} given a realization of $\boldsymbol{Y}_{Jt} = (Y_{2t}, \ldots, Y_{J+1t})$, which we denote by the lowercase $\boldsymbol{y}_{Jt}$, is

$$Y_{1t}|\boldsymbol{y}_{Jt} \sim N\left(\tilde{\mu}, \tilde{\Sigma}\right),$$

where

$$\tilde{\mu} = \sum_{j=2}^{J+1} w_j(\boldsymbol{\lambda}, \sigma) y_{jt},$$

$$\tilde{\Sigma} = 1 + \lambda_1 \sigma^2 (1 - \sum_{j=2}^{J+1} w_j(\boldsymbol{\lambda}, \sigma)\lambda_j), \text{ and}$$

$$w_j(\boldsymbol{\lambda}, \sigma) = \frac{\sigma^2 \lambda_1 \lambda_j}{\sigma_\epsilon^2 + \sum_{j=2}^{J+1} \lambda_j^2 \sigma^2}.$$

Hence, conditional on the realization of the outcomes for the donor units, the distribution of the treated unit depends only on the weights $w_j(\boldsymbol{\lambda}, \sigma)$. We denote the $J \times 1$ vector of such weights by $\tilde{\boldsymbol{w}}$. While the conditions **A1-A2** seem restrictive, the main results in the paper can be extended to include settings with multiple factor loadings and a non-trivial time series component.

2.2.3 Identification and characterization of synthetic controls

In the following proposition we show that $\tilde{\boldsymbol{w}}$ is a minimizer of the statistical risk for the square loss amongst predictors that are linear combinations of the observed outcomes of the donor units.

Theorem 1 (Linear Predictors)**.** *Let $\boldsymbol{Y}_1(0)$ denote the $T_0 \times 1$ vector of outcomes for the treated unit and $\boldsymbol{y}_J$ the $T_0 \times J$ matrix of outcome realizations of the donor units for time periods $1, \ldots, T_0$. Under assumptions **A1-A2** it follows that*

$$\tilde{\boldsymbol{w}} \in \arg\min_{\boldsymbol{w}} \frac{1}{T_0} \mathbb{E}\left[(\boldsymbol{Y}_1(0) - \boldsymbol{y}_J' \boldsymbol{w})' V (\boldsymbol{Y}_1(0) - \boldsymbol{y}_J' \boldsymbol{w})\right],$$

for any positive semi-definite matrix V.

Theorem 1 is simply stating the well known fact that under the square loss the conditional expectation is the best linear predictor. In our case under the linear factor model, the conditional expectation is parametrized by the $\tilde{\boldsymbol{w}}$ weights. While $\tilde{\boldsymbol{w}}$ is a minimizer of the statistical risk amongst linear predictors, it is not immediate whether it can recover the predictive part of treated outcome in future periods, $\lambda_1 F_t$ for $t > T_0$, when the observations $\boldsymbol{y}_{JT}$ are viewed as random under the linear factor model. The following proposition fleshes out the conditions under which $\boldsymbol{y}_{JT_0+1}' \tilde{\boldsymbol{w}}$ converges to $\lambda_1 F_{T_0+1}$.

Theorem 2 (Predictor convergence)**.** *Given **A1-A2** the following hold:*

1. *There exist no values of λ_j that allow $\boldsymbol{y}_{JT_0+1}' \tilde{\boldsymbol{w}} \overset{m.s.}{\to} \lambda_1 F_{T_0+1}$ as $J \to \infty$.*
2. *If $\frac{1}{\|\boldsymbol{\lambda}_J\|_2^2} \sum_j |\lambda_j| \to 0$ as $J \to \infty$, then $\boldsymbol{y}_{JT_0+1}' \tilde{\boldsymbol{w}} \overset{p}{\to} \lambda_1 F_{T_0+1}$.*

Theorem 2 may look surprising as it seems to provide a negative result, but it is in line with the synthetic controls literature. Statement (1) speaks to the fact that identification and asymptotic results for synthetic control estimators often require conditioning on perfect (or good) pre-treatment fit (Abadie et al. 2010). The result can be overturned in the case in which $\sigma_\epsilon^2/\sigma = 0$, that is, when the noise term is completely dominated by the signal (factor structure). Statement (2) provides conditions for convergence in probability. The condition implies that as $J \to \infty$, $\|\boldsymbol{\lambda}_J\|_2^2 \to \infty$ which implies the condition in Ferman and Pinto (2021) that $\|\boldsymbol{w}_J\|_2^2 \to 0$ given the analytic form of $\tilde{\boldsymbol{w}}_J$. Furthermore, as $\|\boldsymbol{\lambda}_J\|_2^2 \to \infty$ for $J \to \infty$ we also recover the treated unit factor loading:

$$\sum_{j=2}^{J} \tilde{w}_j \lambda_j = \frac{\sigma^2 \lambda_1 \|\boldsymbol{\lambda}_J\|_2^2}{1 + \sigma^2 \|\boldsymbol{\lambda}_J\|_2^2} \to \lambda_1.$$

Intuitively, unless we can distribute the error terms over all units as the donor pool grows, we are not able to get consistency. This intuition is similar to the requirements in Ferman and Pinto (2021) and can be thought as justifying their results in our conditional normal setting.

Next, we consider the question of whether $\tilde{\boldsymbol{w}}$ is a *synthetic control*. By this we mean whether the sum to one and non-negative constraints can be justified under our factor model. In general, if λ_1 is fixed and does not depend on other factor loadings then the result will be negative. However, if we allow λ_1 to depend on the $\boldsymbol{\lambda}_J$ then the consistency conditions and synthetic control constraints can be reconciled.

Theorem 3 (Synthetic Control Characterization)**.** *For fixed J under **A1-A2**, $\tilde{\boldsymbol{w}} \in \Delta^J$ iff the following conditions hold*

1. *$sign(\lambda_1) = sign(\lambda_j)$ for all j,*
2. *$\sum_j \lambda_j^2 - \lambda_1 \sum_j \lambda_j + \frac{\sigma_\epsilon^2}{\sigma^2} = 0$.*

Furthermore, the following statements follow

1. *For a fixed λ_1, a sufficient condition for the existence of sequences $\{\lambda_j\}$ such that (1) and (2) hold is that $\lambda_1^2 \geq \frac{4\sigma_\epsilon^2}{J\sigma^2}$.*
2. *For a fixed λ_1, as $J \to \infty$ if $\frac{1}{\|\boldsymbol{\lambda}_J\|_2^2} \sum_j |\lambda_j| \to 0$ then there exist no sequences $\{\lambda_j\}$ for which (2) and (1) hold simultaneously.*
3. *Suppose condition (1) holds, let $\lambda_1 = h(\boldsymbol{\lambda}_J)$ for a component-wise weakly increasing odd function $h : \mathbb{R}^J \to \mathbb{R}$, then if as $J \to \infty$, $\|\boldsymbol{\lambda}_J\|_2^2 \to \infty$, a sufficient condition for $\tilde{\boldsymbol{w}} \in \Delta^J$ is $|h(\boldsymbol{\lambda}_J)| \frac{\|\boldsymbol{\lambda}_J\|_1}{\|\boldsymbol{\lambda}_J\|_2^2} \to 1$.*
4. *For any $\boldsymbol{\lambda}_J$, if $\frac{\|\boldsymbol{\lambda}_J\|_1}{\|\boldsymbol{\lambda}_J\|_2^2}$ is in the convex hull of the entries of $\boldsymbol{\lambda}_J$, then any function h such that $h(\boldsymbol{\lambda}_J) = \boldsymbol{\lambda}_J' \boldsymbol{w}$ for $\boldsymbol{w} \in \Delta_J$ satisfies the condition in (3).*

Theorem 3 clarifies the conditions on the factor loadings under which the target parameter $\tilde{\boldsymbol{w}}$ is a synthetic control. The main result is that a sufficient condition for $\tilde{\boldsymbol{w}}$ to be a synthetic control is

$$|h(\boldsymbol{\lambda}_J)| \frac{\|\boldsymbol{\lambda}_J\|_1}{\|\boldsymbol{\lambda}_J\|_2^2} \to 1,$$

when $\lambda_1 = h(\boldsymbol{\lambda}_J)$. Together with our sufficient conditions from Theorem 2, this implies that λ_1 has to grow with the factor loadings of the donor units. In particular, this rules out the possibility that λ_1 is a fixed constant. This however does not imply that synthetic controls are not possible, in fact, this condition is more general than the one required by Ferman and Pinto (2021). When $\frac{\|\boldsymbol{\lambda}_J\|_1}{\|\boldsymbol{\lambda}_J\|_2^2}$ is in the convex hull of the entries of $\boldsymbol{\lambda}_J$, it follows that there exists a $\boldsymbol{w}^*$ such that $\lambda_1 = \boldsymbol{\lambda}_J' \boldsymbol{w}^*$ that satisfies the condition and, therefore, implies that as

$J \to \infty$, $\tilde{\boldsymbol{w}} \in \Delta^J$. Hence, in settings in which the treated unit factor loading is in the convex hull of the donor units factor loadings the target parameter *is* a synthetic control.

2.2.4 Inference

In this section we consider how to estimate $\tilde{\boldsymbol{w}}_J$ using a data set of pre-treatment outcomes $\{y_{1t}(0), \boldsymbol{y}_{Jt}(0)\}_{t=1}^{T_0}$. Given that we are interested in comparing frequentist and Bayesian procedures we focus on the maximum likelihood estimator. We do not directly observe the factor loadings, but we can estimate the $\tilde{\boldsymbol{w}}$ weights by maximizing the following *pseudo* log-likelihood for parameter $\boldsymbol{\theta} = (\boldsymbol{w}, \Sigma)$:

$$l_{T_0}(\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi\Sigma) - \frac{1}{T_0}\sum_{t=1}^{T_0}\frac{1}{2\Sigma}\left(y_{1t} - \sum_{j=2}^{J+1} w_j y_{jt}\right)^2.$$

We derive our theoretical results for the MLE in two parts. First, in Theorem 4 we show that for fixed J, the size of the donor pool, the MLE can recover the predictive part of the treated unit factor model and has the standard Gaussian approximation as $T_0 \to \infty$. Recall, however, that our characterization and identification results in the previous section required $J \to \infty$. Therefore, we also derive conditions under which as J and T_0 go to ∞ the MLE uniformly converges to the predictive part of the treated unit factor model. The second set of results Theorem 5 and Corollary 6 extend results in the semi-parametric estimation literature to the synthetic control framework.

Theorem 4 (MLE for fixed J)**.** *Let $\hat{\boldsymbol{\theta}}_{MLE} \in argmax_{\boldsymbol{\theta}} l_{T_0}(\boldsymbol{\theta} \in \Theta)$ for a compact parameter space Θ, then under **A1-A2**:*

1. *$\hat{\boldsymbol{w}}_{MLE} \xrightarrow{p} \tilde{\boldsymbol{w}}$ as $T_0 \to \infty$ for fixed J.*
2. *$\sqrt{T_0}(\hat{\boldsymbol{w}}_{MLE} - \tilde{\boldsymbol{w}}) \overset{a}{\sim} N(0, V_{T_0})$ as $T_0 \to \infty$ for fixed J, for $V_{T_0} = \frac{1}{T_0}\mathbb{E}[(\nabla_{\boldsymbol{w}} l_{T_0}(\tilde{\boldsymbol{\theta}})\nabla_{\boldsymbol{w}} l_{T_0}(\tilde{\boldsymbol{\theta}})')^{-1}]$, where $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{w}}, \tilde{\Sigma})$.*

For fixed J, the MLE consistency and asymptotic normality result is straightforward. Theorem 4 shows that the target parameter $\tilde{\boldsymbol{w}}$ can be consistently estimated by the MLE. However, as discussed, our identification result requires that the donor pool grows with the sample size, $J \to \infty$. This means that the parameter space is growing with the sample size T_0. In order to account for this, Theorem 5 provides conditions for uniform consistency and asymptotic normality to the target parameter as $J, T_0 \to \infty$.

Theorem 5 (MLE with growing J)**.** *Let $\hat{\boldsymbol{\theta}}_{MLE} \in argmax_{\boldsymbol{\theta}} l_{T_0}(\boldsymbol{\theta} \in \Theta)$ for a compact parameter space Θ, then under **A1-A2** and λ_j are uniformly bounded:*

1. $\frac{1}{T_0}\sum_t \boldsymbol{y}_{Jt}\boldsymbol{y}'_{Jt} = D_{T_0}$ *where* $0 < lim\ inf_{T_0}\sigma_{min}(D_{T_0}) \le lim\ sup_{T_0}\sigma_{max}(D_{T_0}) < \infty$,

2. $max_{t\le T_0}\|\boldsymbol{y}_{Jt}\|_2^2 = O_p(J)$,

3. $\sup_{\boldsymbol{\beta},\boldsymbol{\gamma}\in\mathcal{S}_J(1)}\sum_t |\boldsymbol{y}'_{Jt}\boldsymbol{\beta}|^2|\boldsymbol{y}'_{Jt}\boldsymbol{\gamma}|^2 = O_p(T_0)$.

Then, it follows that if $o(T_0) = J(\log J)^3$

$$\|\hat{\boldsymbol{w}}_{MLE} - \tilde{\boldsymbol{w}}\|_2^2 = O_p(J/T_0).$$

If $o(T_0) = J^2\log(J)$ *then*

$$\sqrt{T_0}\boldsymbol{\alpha}'(\hat{\boldsymbol{w}}_{MLE} - \tilde{\boldsymbol{w}})/\sigma_\alpha \xrightarrow{d} N(0,1),$$

for any $\boldsymbol{\alpha} \in S_J(1)$ *and*

$$\sigma_\alpha^2 = (\mathbb{E}[\epsilon_{Jt}^2])\boldsymbol{\alpha}'D_{T_0}^{-1}\boldsymbol{\alpha},$$

where $S_J(1)$ *denotes the Euclidean ball of radius 1 in* $\mathbb{R}^J$.

Theorem 5 shows that L^2 norm convergence and uniform Gaussian approximation is possible when T_0 grows at rate faster J. In particular, we require that T_0 grows faster than J for the consistency result and that T_0 grows faster than J^2 for the asymptotic normality result. While these rates might be impractical in settings with small T_0, they speak to the discussion in Abadie et al. (2010) and Abadie and Vives-i-Bastida (2025) that large donor pools may increase the bias in synthetic control estimators. Intuitively, larger donor pools imply more parameters to estimate which might increase the finite sample bias of the estimator. Next, we apply this result to our specific setting when we use the data $\boldsymbol{y}_{JT_0+1}$ to predict the treated unit outcome in absence of the intervention.

Corollary 6. *Under the conditions of Theorem 5, as* $J, T_0 \to \infty$*:*

1. *If* $o(T_0) = J(\log J)^3$ *and* $\frac{1}{\|\boldsymbol{\lambda}_J\|_2^2}\sum_j |\lambda_j| \to 0$, *then*

$$\boldsymbol{y}'_{JT_0+1}\hat{\boldsymbol{w}}_{MLE} \xrightarrow{p} \lambda_1 F_{T_0+1}.$$

2. *If* $o(T_0) = J^2\log(J)$ *and* $\frac{1}{\|\boldsymbol{\lambda}_J\|_2^2}\sum_j |\lambda_j| \to 0$, *then*

$$\sqrt{T_0}(\boldsymbol{y}'_{JT_0+1}\hat{\boldsymbol{w}}_{MLE} - \lambda_1 F_{T_0+1})/\sigma_{y_{JT_0+1}} \xrightarrow{d} N(0,1).$$

Corollary 6 provides conditions for valid frequentist inference to the predictive part of the treated unit factor model as $T_0, J \to \infty$. Similar semi-parametric results have also been derived in Ferman (2021). Our results are different in that they apply to a wider class of models under our characterization conditions and provide explicit rate conditions for J and T_0. Indeed, all conditions are imposed directly on the factor structure.

2.3 The Bayesian Synthetic Control

We propose the Bayesian equivalent of the program described in section 2 to generate synthetic controls. In particular, our Bayesian formulation includes two key aspects of synthetic controls: (1) the synthetic treated unit is a convex combination of the donor units, that is, we do not want to extrapolate outside the convex hull of the donor units and (2) we construct the synthetic control by matching the predictors of treated unit and donor units for the outcome of interest. An advantage of the Bayesian approach is that we can directly quantify the uncertainty in our estimates. In this section, first we discuss a valid Bayesian inference procedure and then the Bayesian synthetic control model.

2.3.1 Bayesian Inference

We derive conditions for valid Bayesian inference in a general setting that includes our set up. Following Imbens and Rubin (1997) and Pang et al. (2022) we consider the Bayesian inference problem as a missing data problem. Let the adoption process of the intervention for one treated unit be encoded by a treatment random variable $\mathbf{d}_1 = (d_{11}, \ldots, d_{1T})$ that is fully determined by an adoption time random variable a_1. Similarly, we can define these variables for the donor pool as $\mathbf{D}$ and $\mathbf{a}$.

The potential outcome function of our outcome of interest Y_{it} under SUTVA is given by $Y_{it}(\mathbf{d}_i(a_i)) = Y_{it}(a_i)$, where no anticipation implies that $Y_{it}(a_i) = Y_{it}(c)$ for $t < a_i$ and c denoting the counterfactual state in which the unit is never treated.

Assumptions:

1. **(B1) - (B2)**: SUTVA and no anticipation.

2. **(B3) Latent Ignorability**: there exist latent variables $\mathbf{U}_i = (u_{i1}, \ldots, u_{iT})$ such that:

$$\mathbf{y}_i(0) \perp \mathbf{D}_i | \mathbf{U}_i.$$

3. **(B4) Exchangeability**: given $\mathbf{U}$, permutations of the indices it of the sequence $\{(Y_{it}(0)\}_{i\in[N],t\in[T}$ do not alter the joint distribution.

The above assumptions allow us to get an expression for the posterior on the "missing" data, the outcome counterfactual for treated units after treatment assignment if they had not been treated. The predictive inference consists in sampling from the posterior distribution of the missing data conditional on the observed data and a model indexed by parameter vector $\boldsymbol{\theta}$ with prior distribution $\pi(\theta)$. Let the missing and observed data be denoted by $\mathbf{Y}^{mis} = \{Y_{1t}\}_{t>T_0}$ and $\mathbf{Y}^{obs} = (\{Y_{it}\}_{i\in[J+1],t\leq T_0}, \{Y_{jt}\}_{j\in[J],t>T_0}, \mathbf{D})$. Then, under **B1-B3** we can factor out the assignment:

$$\begin{aligned}
P(\mathbf{Y}^{mis}|\mathbf{Y}(0)^{obs}, \mathbf{D}, \theta) &\propto P(\mathbf{Y}(0)^{mis}, \mathbf{Y}(0)^{obs}, \mathbf{D}, \theta) \\
&\propto P(\mathbf{Y}(0), \mathbf{U})P(\mathbf{D}|\mathbf{Y}(0), \theta) \\
&\propto P(\mathbf{Y}(0), \theta)P(\mathbf{D}|\theta) \\
&\propto P(\mathbf{Y}(0), \theta).
\end{aligned}$$

Then under **B4** by de Finetti's Theorem we can separate out the posterior predictive distribution and the likelihood

$$\begin{aligned}
P(\mathbf{Y}(0)^{mis}|\mathbf{Y}(0)^{obs}, \mathbf{D}, \theta) &\propto P(\mathbf{Y}(0), \theta) \\
&\propto \int \Pi_{it} f(Y_{it}(0)|\theta)\pi(\theta)d\theta \\
&\propto \int \underbrace{\left(\Pi_{it\in \text{ mis}} f(Y_{it}(0)^{mis}|\theta)\right)}_{\text{Posterior Predictive Distribution}} \underbrace{\left(\Pi_{it\in \text{ obs}} f(Y_{it}(0)^{obs}|\theta)\right)}_{\text{Likelihood}} \pi(\theta)d\theta
\end{aligned}$$

where f is the marginal density of the potentially observable data.

The assumptions imposed by our sampling model **A1-A2** in which the data is i.i.d and the potential outcomes are given by a latent factor model implicitly satisfy **B1-B4**. Therefore, under our sampling we will be able to sample from the Bayes posterior distribution to recover the distribution of the missing data and recover the distribution of the target treatment effect. In order to estimate the model, we use a HMC sampler (NUTS) to get M draws $\{\theta^m\}$ and then our predictive posterior is given by the realizations of

$$\tau_{1t}^m = Y_{1t}(1) - y_{1t}^m(0),$$

In Section 3.5 we describe the Bayesian estimation procedure we implement in the *bsynth* package in more detail. In the next section, we consider under which conditions on the prior model can we use a Bayes estimator to approximate the frequentist inference.

2.3.2 Bayesian Model

A Bayesian model imposes a functional form for the prior distribution and data density f, by modelling explicitly the potential outcome in absence of treatment. In the previous section we showed under which conditions can we perform valid Bayesian inference. In this section, we impose restrictions on the prior structure. We start by considering a Gaussian prior model

$$
\begin{aligned}
y_{1t}|\boldsymbol{y}_{Jt}, \boldsymbol{w}, \sigma_y &\sim N(\boldsymbol{y}'_{Jt}\boldsymbol{w}, \sigma_y^2), \\
w_j|\boldsymbol{y}_{Jt} &\sim N(\mu_j, \tau_j^2).
\end{aligned}
$$

Given that the Gaussian conjugate prior is Gaussian, it can be shown that the **Bayes estimator** for the implicit weights is given by

$$
\hat{w}_j^B = \mathbb{E}_B[w_j|\boldsymbol{y}_t] = \int w_j p(w_j|\boldsymbol{y}_t) dw_j.
$$

Furthermore, in this case the predictive posterior distribution is normally distributed and

$$
\begin{aligned}
\hat{Y}_{1t}^B = \boldsymbol{y}'_{Jt}\mathbb{E}_B[w_j|\boldsymbol{y}_t] &= \frac{\sigma_y^2}{\sigma_y^2 + \sum_j \tau_j^2}\boldsymbol{y}'_{Jt}\mu_J + \frac{\sum_j \tau_j^2}{\sigma_y^2 + \sum_j \tau_j^2} y_{1t}, \\
\mathbb{V}_B(\boldsymbol{y}'_{Jt}\boldsymbol{w}|\boldsymbol{y}_t) &= \frac{\sigma_y^2 \sum_j \tau_j^2}{\sigma_y^2 + \sum_j \tau_j^2}.
\end{aligned}
$$

Motivated by the characterization of synthetic controls in Theorem 3 we consider Bayesian models in which the parameters μ_j are in the simplex. This suggests adding the following restriction:

$$
\mu_j \sim Dir(1),
$$

where $Dir(1)$ denotes the Dirichlet distribution with scale one. Intuitively, this restriction forces the means of the target weights to be in the simplex.

In Section 4 we describe alternative Bayesian models implemented in the *bsynth* package that allow for additional covariates and a Gaussian process term. The Bayesian synthetic control with additional covariates, which has not been implemented in the literature yet, and the use of Gaussian processes, also suggested by Arbour et al. (2021). Our theoretical results could potentially be extended to apply to both cases with additional regularity conditions. In the case of the Gaussian process note that in a general setting a BvM result for Gaussian processes was derived in Ray and van der Vaart (2020). In the next section, we link the Bayesian and frequentist inference for the baseline Bayesian model.

2.3.3 Bernstein-von Mises Result

In this section we consider how the Bayesian inference can be used to approximate the frequentist inference. We derive a Bernstein-von Mises style result in which under the correct prior specification the Bayesian posterior predictive distribution converges in the total variation sense to the MLE sample distribution as $T_0, J \to \infty$. Intuitively, our result states that if we assume that the factor loading of the treated unit can be recovered by a convex combination of the treated units then under the same assumptions that yield a valid MLE estimator, the Bayes estimator is able to consistently estimate the predictive term $\lambda_1 F_{T_0+1}$. If additionally, the uncertainty in our predictions converges to the frequentist sampling variance, then the two estimators distributions are close in the total variation sense.

Theorem 7 (BvM)**.** *Under **A1-A2**, the assumptions of Corollary 6 and*

1. ***Prior conditions:*** $\|\mu_J\|_2^2 \to 0$, $\{\tau_j\}$ *such that* $\sum_j \tau_j^2 = O(J^\alpha)$, *for* $0 < \alpha < 1$, *as* $J \to \infty$, *and* $\sigma_y \to \sigma_\epsilon$.

2. ***Convex recovery:*** $\|\lambda_1 - \boldsymbol{\lambda}_{\boldsymbol{J}}' \boldsymbol{\mu}_J\|_2 \to 0$ *as* $J \to \infty$.

Then, as $T_0, J \to \infty$ *at rate* $o(T_0) = J^2 \log(J)$,

$$\boldsymbol{y}'_{JT_0+1} \mathbb{E}_B[\boldsymbol{w} | \boldsymbol{y}_{T_0}] \xrightarrow{p} \lambda_1 F_{T_0+1},$$

and

$$\|\Phi^{MLE}_{T_0,J} - Q_{T_0,J}\|_{TV} \to 0,$$

where $\Phi^{MLE}_{T_0+1,J}$ *denotes the MLE finite sample distribution and* $Q_{T_0+1,J}$ *the Bayes posterior predictive distribution.*

Theorem 7 imposes strong conditions on the class of priors necessary to approximate the MLE distribution. In particular, it is key that we choose the μ_j in the simplex and in a way that recovers the treated unit factor loading. The requirement that such a sequence of priors exists is motivated by our characterization conditions in Theorem 3. In the same spirit as in the frequentist synthetic control, if the Bayesian synthetic control is unable to find implicit weights that replicate the treated unit outcomes in the pre-treatment period, then it is likely the Bayesian synthetic control will be biased as the model is miss-specified.

The other requirement in the proof of Theorem 7 is that the Bayesian posterior is Gaussian. While this requirement is important for the proof method, which relies on the analytical form of the KL divergence between Gaussian distributions, it is not a necessary requirement. A more general result could be derived by imposing weaker restrictions on the functional form and the second moments of the posterior distribution.

The BvM result in Theorem 7 is important because it gives conditions under which researchers might be able to interpret their Bayesian credible intervals as valid confidence intervals. As $T_0, J \to \infty$

under the conditions of Theorem 7 the $1-\alpha$ credible interval defined by the limit Bayesian posterior predictive distribution and the $1-\alpha$ confidence interval defined by the MLE estimator will coincide. Hence, using Bayesian synthetic controls offers a new way of performing valid asymptotic inference for synthetic controls without the need of exact inference or permutation tests.

2.3.4 Simulation Evidence

In this section we compare the standard synthetic control and the Bayesian synthetic control for a grouped linear factor model data generating process. Similar data generating processes have been used to study the properties of synthetic controls estimators, for example in Firpo and Possebom (2018) or in Abadie and Vives-i-Bastida (2025). In particular, we let the potential outcome in absence of intervention be given by

$$Y_{it}(0) = \lambda_{f(i)t} + \epsilon_{it}. \tag{2.1}$$

where the λ_{ft} follow an $AR(1)$ with $\rho = 0.5$, standard Gaussian innovations and noise given by $\epsilon_{it} \sim N(0, \sigma^2)$. Without loss of generality we assume that only unit 1 is treated and that the treatment effect is zero. We consider two designs: a *dense* design in which units are grouped in 2 groups of 10 units, such that $f(1) = f(2) = \cdots = f(10)$, and a *sparse* design in which units are paired in 10 groups, such that $f(1) = f(2)$. In the *dense* case the optimal synthetic control assigns equal weight to units 2 to 10, whereas in the *sparse* case the optimal synthetic control puts all the weight on unit 2.

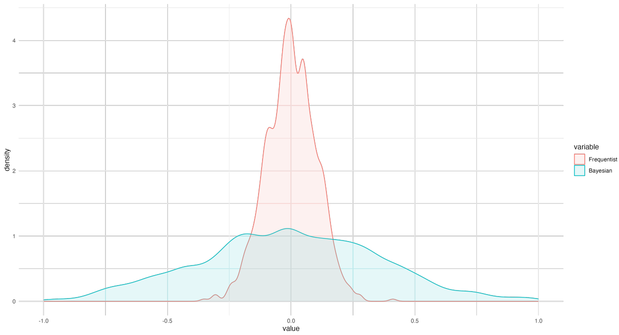
The simulation settings satisfy the convex recovery assumptions of Theorem 3 and Theorem 7 as there exists a sequence of weights that perfectly recreates the factor loading of the treated unit. In the sparse case, for example, $w_2 = 1$ and $w_j = 0$ for all $j > 2$. However, each setting does not necessarily satisfy the *density* condition, or the prior condition, meaning that the target parameter does not satisfy that $\|\tilde{\boldsymbol{w}}\| \to 0$ as $J \to \infty$. In fact, to check the relevance of the proposed Bayesian method and the theoretical results we consider a realistic case with a fixed number of units $J = 20$. We will see however, that in this realistic setting, there will still be BvM convergence in both the *dense* and *sparse* designs as $T_0 \to \infty$.

In our simulation analysis, for each T_0 we estimate the standard synthetic control and report our estimated treatment effect on the treated for 10000 draws. We compute the treatment effect over 10 post-treatment periods $\hat{\tau}_1 = \frac{1}{T-T_0}\sum_t \hat{\tau}_{1t}$, which allows us to average over the additional noise terms ϵ_{it} for $t > T_0$. We compare this empirical distribution to the mean posterior predictive distribution of our Bayesian model for the 10000 draws. Figure 2.1 shows the empirical distribution for each estimator in the dense setting for different values of T_0.

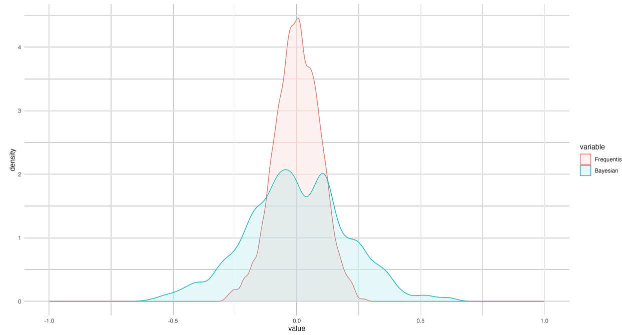
As can be seen in Figure 2.1 the Bayesian estimator exhibits over coverage when T_0 is small. However, for moderate values of T_0, such as 40 or 60 time periods, the coverage is close to that of the frequentist estimator. As T_0 grows, the prior becomes dominated, and the Bayesian posterior distribution converges to the finite sample frequentist distribution, as can be seen in panels (e) and (f) in Figure 2.1. A similar pattern can be observed in Figure 2.2 for the *sparse* design, albeit, as expected, the convergence rate is slower than in the dense case. Panel (e) in Figure 2.2 shows that for large values of T_0 the empirical CDFs of the two estimators coincide, and panel (f) shows that the empirical total variation distance between the two distributions decreases as T_0 grows. Overall, the main takeaway is that in both sparse and dense settings for medium sized panels the Bayesian posterior distribution can be used to approximate the frequentist finite sample distribution.

Consequently, for applied researchers with medium T_0 and J it may be possible to use Bayesian synthetic control methods and interpret the Bayesian credible intervals for the estimated treatment effects as valid confidence intervals, providing a new way of doing inference in the framework of synthetic controls.

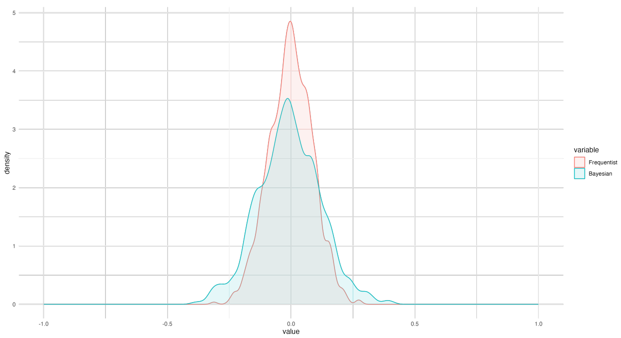
Figure 2.1: Convergence of frequentist and Bayesian coverage as $T \to \infty$ for the dense case.



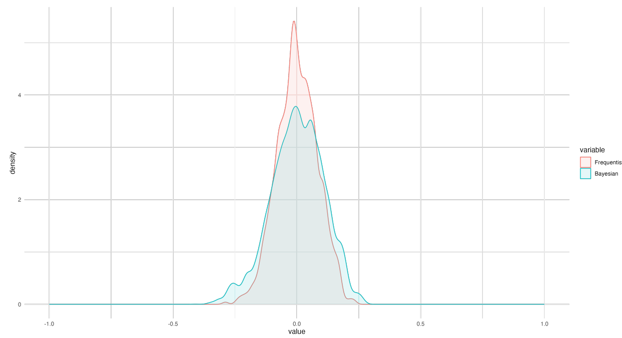
(a) $T_0 = 10$



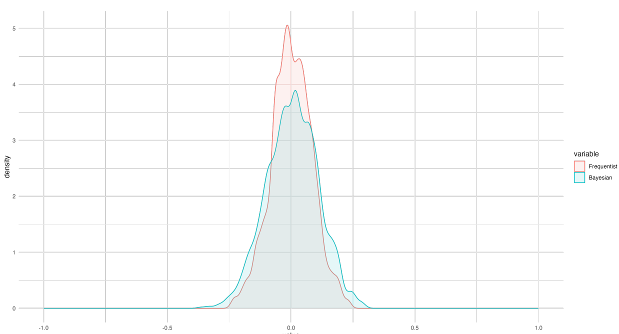
(b) $T_0 = 20$



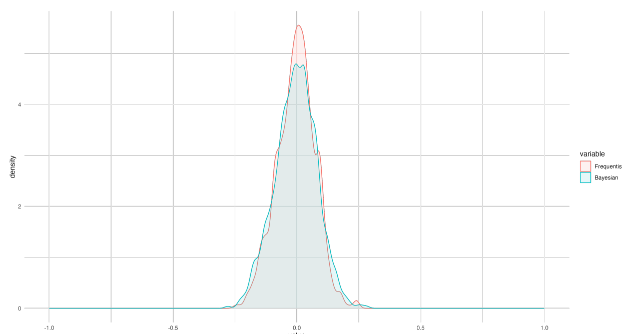
(c) $T_0 = 40$



(d) $T_0 = 60$



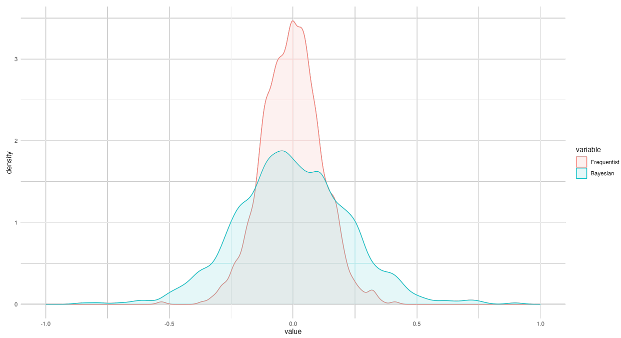
(e) $T_0 = 80$



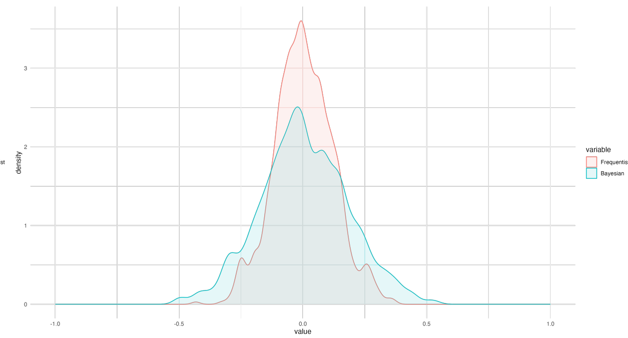
(f) $T_0 = 200$

Notes: Kernel densities of the frequentist empirical distribution of the estimated treatment effect (in red) and the Bayesian posterior mean (in blue) over 10000 draws for different values of T_0. The potential outcomes are generated by the grouped factor model (2.1) with $\sigma = 0.25$ in a dense design with factor loadings grouped in two groups of 10 units.

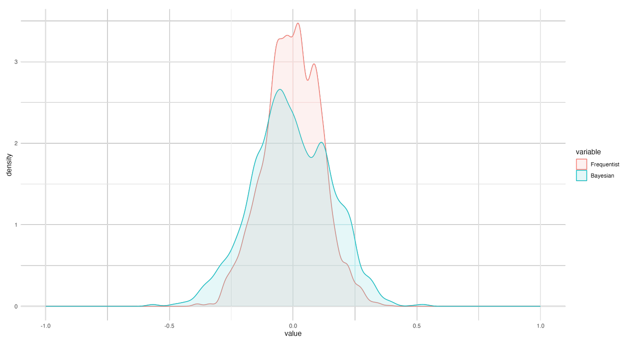
Figure 2.2: Convergence of frequentist and Bayesian coverage as $T \to \infty$ for the sparse case.



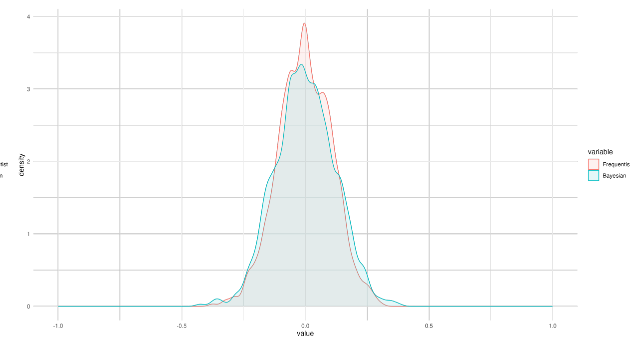
(a) $T_0 = 40$



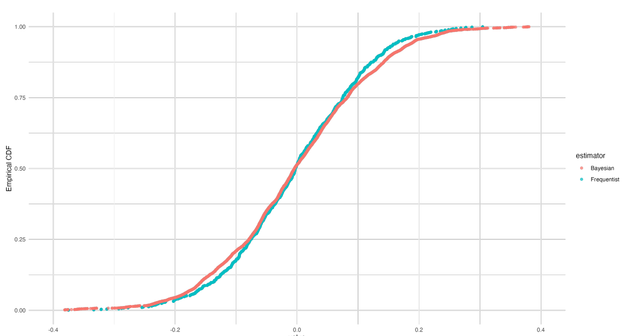
(b) $T_0 = 60$



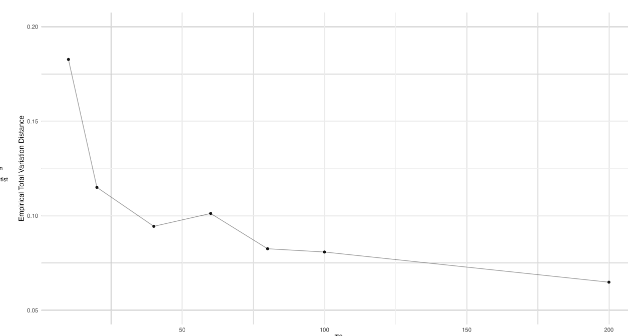
(c) $T_0 = 80$



(d) $T_0 = 200$



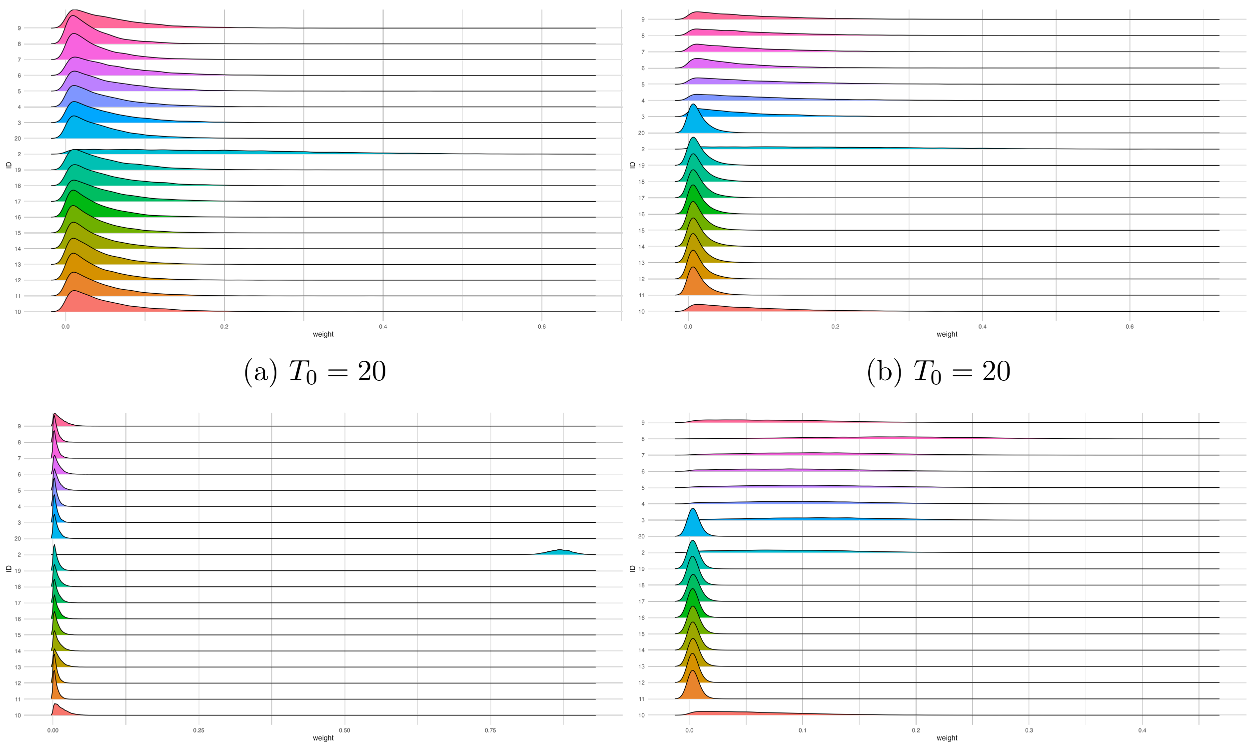
(e) CDF for $T_0 = 200$



(f) Total variation distance

Notes: Kernel densities of the frequentist empirical distribution of the estimated treatment effect (in red) and the Bayesian posterior mean over 10000 draws (in blue) for different values of T_0. Panel (e) shows the empirical CDF for each case. Panel (f) shows the empirical total variation distance between both distributions for different values of T_0. The potential outcomes are generated by the grouped factor model (2.1) with $\sigma = 0.25$ in a sparse design with factor loadings grouped in two groups of 10 units.

Figure 2.3: Implicit weights as $T_0 \to \infty$.



(a) $T_0 = 20$

(b) $T_0 = 20$

(c) $T_0 = 200$

(d) $T_0 = 200$

Notes: Kernel densities of the Bayesian implicit weights for different values of T_0. The potential outcomes are generated by the grouped factor model (2.1) with $\sigma = 0.25$. Panel (a) and (c) refer to the *sparse* design and panels (b) and (d) refer to the *dense* design.

Another point highlighted by the simulations is the use of Bayesian synthetic controls in consistently estimating the synthetic control weights. Figure 2.3 shows the implicit weights posterior distribution of the Bayesian estimator for different values of T_0. As can be seen in panels (a) and (c) for the sparse design and (b) and (d) for the dense design, as $T_0 \to \infty$, the mean of the implicit weight distribution is centered around the optimal value of the synthetic controls ($\tilde{w}$s), which in the sparse case is $\tilde{w}_2 = 1$ and in the dense case is $\tilde{w}_2 = \cdots = \tilde{w}_{10} = 1/9$. This is evidence of unit wise convergence which implies the L^2 uniform convergence result in Theorem 5. Overall, this adds to the previous discussion that Bayesian synthetic control methods may be useful both for performing inference for the estimated treatment effect and in interpreting the synthetic control weights.

In the appendix, we also consider simulations in which we compare the 10000 frequentist draws to the posterior distribution of the Bayesian model for 1 draw and show that even in this case as $T_0 \to \infty$ we have convergence between the two distributions.

2.3.5 The *bsynth* package

We have implemented the Bayesian synthetic control model we propose in this paper in the publicly available *bsynth* R-package.[1] The *bsynth* R-package extends the Bayesian model we propose to include more complex models. In particular, it allows for additional features such as modelling the time series component with a Gaussian process and adding additional covariates. The *bsynth* package allows for Bayesian models with the following form:

$$\begin{aligned} X_1|w,\sigma &\sim N(X_0 w + f_{1t}, \sigma^2 \mathrm{diag}(\Gamma)^{-2}), \\ w &\sim Dir(1), \\ f_{1t} &\sim \mathcal{GP}, \\ \sigma &\sim N(0,1)^+, \\ \Gamma &\sim Dir(v_1,\dots,v_K), \qquad v_k \in \Delta^k, \end{aligned}$$

where here X_1 and X_0 denote the design matrices for the treated and donor units which may include the outcome variables as well as additional predictors. The f_{1t} term is modelled through a Gaussian process and the weight of the predictors are modelled by Γ. To preserve the main features of synthetic controls both the w and the v weights are assumed to be in the simplex.

The *bsynth* package offers the possibility to compute different statistics of the posterior distribution. Of special interest is an upper bound on the frequentist bias given by the Bayesian model. This bound is motivated by the finite sample bound first developed in Abadie et al. (2010) and expanded in Vives-i-Bastida (2022) to include additional predictors. Given the BvM style result from Theorem 7 the bound can be used to check the likelihood that the Bayesian synthetic control is badly biased due to model mispecification. Intuitively, if the Bayesian synthetic control can not replicate in the pre-treatement period the outcomes of the treated unit then our frequentist interpretation of the method will be biased in the same way standard synthetic controls are biased when perfect pre-treatment fit can not be achieved. In the following section we use the Bayesian synthetic control to study two political economy questions.

2.4 Empirical applications

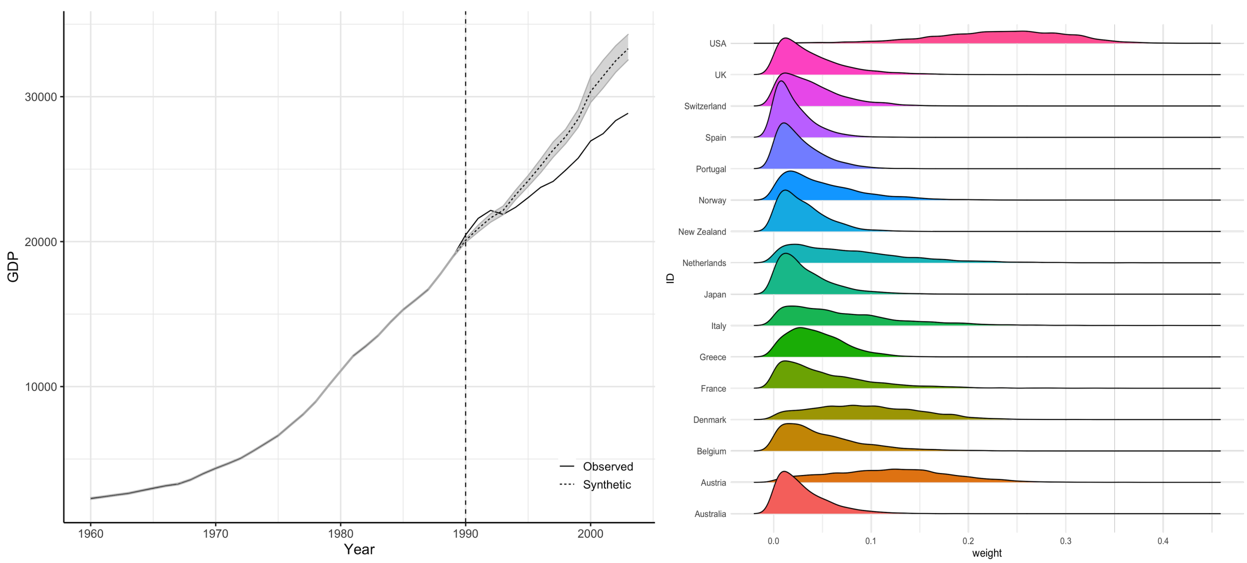
One of the most salient applications of synthetic controls is the study of the impact of the German re-unification in 1990 to the GDP of West Germany. In this paper, we replicate this finding using the Bayesian synthetic control and we highlight the usefulness of the Bayesian inference procedure.

[1] The package can be accessed at https://github.com/google/bsynth.

2.4.1 Re-visiting the German re-unification

In 1989, after the fall of the Berlin wall, the process of re-unifying West Germany and East Germany started. Abadie et al. (2015) found that in absence of the re-unification, West Germany's GDP would have been 8% higher in 2003, 13 years later. Using the same data and specification as Abadie et al. (2015) in Figure 2.4 we display the Bayesian synthetic control for West Germany over and the marginal distribution of implicit weights of the donor units.

Figure 2.4: Bayesian synthetic control for West Germany.



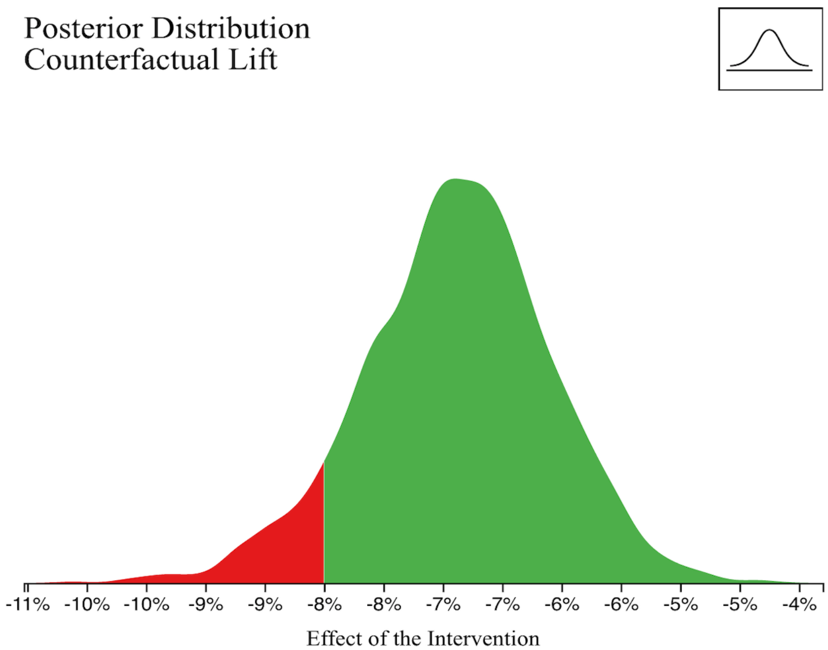
(a) Treatment effect (b) Implicit weight marginals

Notes: Panel (a) shows the West Germany GDP per capita and Bayesian synthetic control estimates from 1960 to 2003 with credible intervals shaded in grey. Panel (b) shows the marginal distributions of the implicit weights of the Bayesian synthetic control.

The Bayesian synthetic control in Figure 2.4 is similar to the standard synthetic control in Abadie et al. (2015). It shows an overall increasing trend in absence of the intervention with a slight reversal in the first few years after the intervention. Similarly, in panel (b) the implicit weights of the Bayesian synthetic control also indicate that West Germany is best replicated by a combination the United States and Austria. In the appendix, we also report the correlations between the implicit weights, a statistic that could be of interest to applied researchers seeking to understand the robustness of their synthetic control estimate. The correlation matrix shows that the implicit weight distributions of the United Sates and Austria are positively correlated, as expected, but it also shows that an alternative synthetic control could have included Denmark, Spain, Portugal and Australia which are also correlated with each other.

An advantage of the Bayesian synthetic control is that we estimate a full posterior distribution. Figure 2.5 shows the treatment posterior distribution for the post-treatment period relative to the baseline year. The mean of the posterior distribution is 7.5% which is slightly lower than the frequentist estimate of the ATET of 8%. The 95% Bayesian credible interval spans $[-9\%, -6\%]$ and contains the frequentist estimate. Given that in this setting the number of units is small ($J = 16$) and the number of pre-treatment periods is moderate ($T_0 = 30$), we expect the Bayesian credible interval to be a good approximation of the confidence interval. Furthermore, in the appendix, we show the posterior distribution of the frequentist bias bound that depends on the MAD in the pre-treatment period. The bound indicates that the likelihood that the sign of the result could be overturned due to the bias induced by model misspecification (bad pre-treatment fit) is small.

Figure 2.5: Treatment effect posterior distribution.



Notes: Posterior distribution of the average treatment effect in the post-treatment period, the change of color indicates the frequentist estimated effect of 8%. This figure is generated directly from the output of the *bsynth* R-package.

2.4.2 The impact of the Catalan secession movement

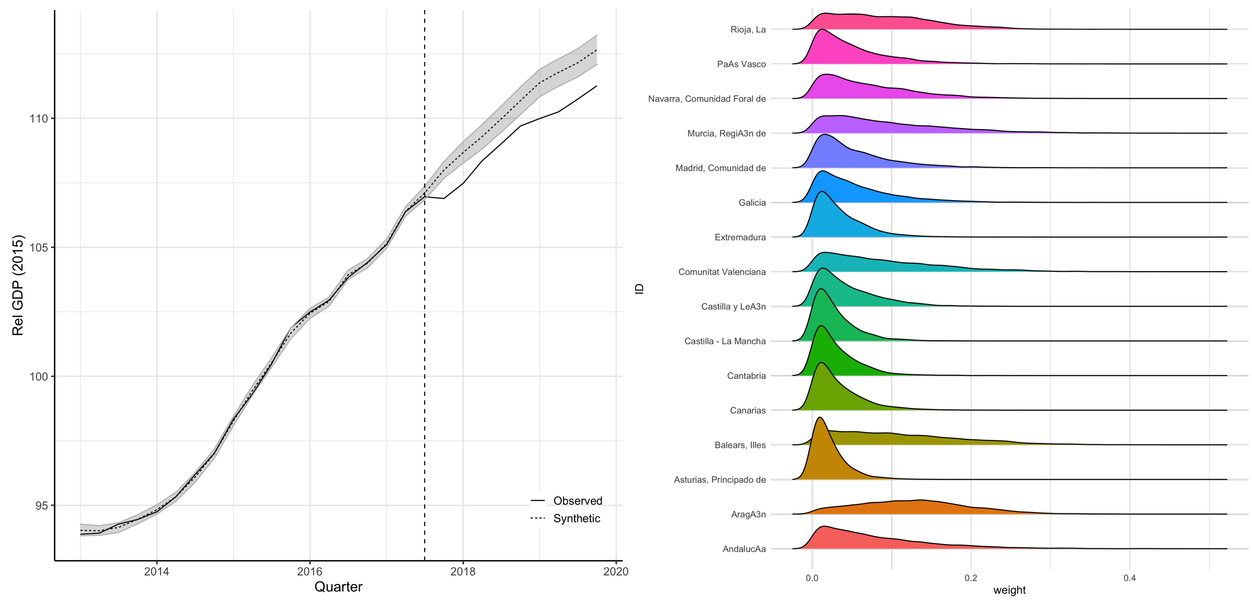
The pro-independence demands of the Catalan electorate, commonly referred as the Catalan secession process, have been the focus of Catalan politics, and Spanish politics, since 2012. While the movement started with a plea for higher fiscal autonomy (the ability of the Catalan government to collect and administer its own taxes), quickly it shifted towards independence through a series of plebiscitary elections that were won by pro-independence parties. This transition towards a pro-independence mandate culminated in 2017 with the celebration of a non-binding independence referendum on the 1st of October which received wide international media coverage. Following the referendum, on the 27th of October a unilateral declaration of independence (UDI) was voted and approved by the

Catalan parliament. However, the UDI was deemed unconstitutional by the Spanish constitutional court and the Spanish government suspended the Catalan government (activating the *article 155*). More so, the members of the Catalan parliament and the government behind the UDI were arrested while several high ranking members of the government, including the president Carles Puigdemont, sought asylum in other countries.

Overall the UDI and the events that ensued after the Spanish government suspended the Catalan government had a great impact on the Catalan economy. An extensive analysis of different outcomes affected by the UDI has been done by Esteller-Moré and Rizzo (2022), also using a synthetic control methodology. The main finding in Esteller-Moré and Rizzo (2022) is that the UDI lead to a significant drop in short term bank deposits (capital out-flow), accounting to up to 2.5bn euros. This result is in line with reported evidence in the media[2] that the UDI and political instability lead to many firms and individuals to reallocate outside of Catalonia. In this paper, we focus on quarterly GDP as our main outcome of interest. To estimate the Bayesian synthetic control we consider all quarters between 2012 and the UDI in 2017 as the pre-treatment period and all quarters from the UDI to 2020 as the post-treatment period. We choose this time frame to avoid the effects of the 2008 financial crisis and the 2010 European debt crisis. Given that Catalonia has a different economic structure than other regions in Spain the effect of the two crises was different and therefore matching outcomes in that time period is potentially misleading. We also restrict the post period to avoid the covid-19 shock in 2020.

[2]See for example the following Financial Times article: https://www.ft.com/content/18f0ca8c-607b-4633-a83f-e99f71a046e8.

Figure 2.6: Bayesian synthetic control for Catalonia.



(a) Treatment effect

(b) Implicit weight marginals

Notes: Panel (a) shows the Catalan quarterly GDP (relative to 2015) and Bayesian synthetic control estimates from 2013 to 2020 with 75% credible intervals shaded in grey. Panel (b) shows the marginal distributions of the implicit weights of the Bayesian synthetic control.

Figure 2.6 shows the Bayesian synthetic control and implicit weight marginal distributions for Catalonia. The main finding is that the UDI lead to a quarterly GDP decrease of 1% over 2017-2020 relative to the third quarter of 2017. The Bayesian synthetic control allows us to quantify uncertainty with the 75% credible interval being $[-0.6\%, -1.3\%]$ and the 95% credible interval slightly larger at $[-0.3\%, -1.6\%]$. Overall this is indicative evidence that the capital flight described in the press and Esteller-Moré and Rizzo (2022) may have had a significant impact on the Catalan GDP. In panel (b) of Figure 2.6 we can see that the Bayesian synthetic Catalonia is mostly composed of neighboring regions: Aragon, the Balearic islands and Valencia. A potential concern is that the UDI also affected positively the regions that received the out-flowing capital, biasing the estimated treatment effect upwards. In particular, as detailed in media coverage, many Catalan firms migrated their fiscal head quarters to Madrid or the neighboring Valencia.[3] In light of this, in the appendix, we re-do the exercise removing Madrid and Valencia from the donor pool and find similar treatment effects.

2.5 Conclusion

This paper contributes to the synthetic control literature in two ways. First, we characterize the conditions on the primitives of factor models (the factor loadings) that generate target parameters

[3] See the article in footnote 2 for reference.

(minimizers of the statistical risk) that are *synthetic controls*. This result complements the existing literature on the asymptotic properties of synthetic controls under linear factor models by providing guidance on the set of data generating processes for which synthetic control estimators are best suited. We show that the target parameters can be estimated by MLE and derive rate conditions for uniform consistency as the number of time periods and donor units grow.

Second, we propose the Bayesian synthetic control as an alternative to perform inference for synthetic control methods. We derive a Bernstein-von Mises style result that states conditions under which the Bayesian synthetic control and the MLE estimator converge in the total variation sense. This result can be used to approximate the frequentist inference using the Bayesian synthetic control in large samples. Through simulations we show that this result might also be useful in settings in which the theoretical assumptions don't hold, highlighting the potential for theoretical extensions in future work.

Finally we apply the Bayesian synthetic control to study two important political economy interventions: that of the re-unification of a country and that of secession. We replicate using the Bayesian synthetic control the findings in Abadie et al. (2015) that the German re-unification lead to a significant decrease in West Germany's GDP and show that the Catalan unilateral declaration of independence of 2017 may have lead to a 1% decrease in GDP. The Bayesian synthetic control allows us to quantify uncertainty and in both applications we can conclude that the effects are unlikely to be zero.

Future work may aim to generalize the results in the paper to a larger class of models and extend the functionality of the *bsynth* R-package.

Chapter 3

Predictor Selection for Synthetic Controls

3.1 Introduction

Synthetic controls have become a popular method for making inferences on causal effects of policy interventions. Its applications have ranged from evaluating the effects of tax changes (Abadie et al. (2010), Kleven et al. (2013)) to more complex policy changes such as guaranteed job programs (Kasy and Lehner (2021)). Beyond the social sciences synthetic controls have also been used in other applied sciences, for example in engineering and the health sciences, and in the private sector, for instance to evaluate the effect of advertising promotions in large tech companies.

In the classical synthetic control setting (Abadie and Gardeazabal (2003)) a single aggregate unit (such as a city or a state) is exposed to a policy treatment at period T_0, and J units that are never exposed to the policy (called the donor units) are used to generate a counterfactual. The researcher builds the synthetic control by finding the combination of donor units that best matches the pre-treatment characteristics (called the predictors) of the treated unit. Predictors may include both pre-treament outcomes as well as other covariates that are informative of the outcome of interest. An open question in the literature regards best practices and theory on how to choose the set of predictors and how to weight the importance of each predictor in building the synthetic control.

There are two main reasons why the choice of predictor weights matters for computing synthetic controls. First, as noted by Abadie et al. (2015), Abadie and Vives-i-Bastida (2025), Klößner et al. (2018) and Malo et al. (2020) among others, the predictor weights are important for the performance of the synthetic controls. Weighting the predictors naively can lead to bad pre-intervention fit and synthetic controls with poor post-intervention performance. This, in turn, means that the estimated treatment effects will be biased. Second, the predictor weights may be of interest in their own right. The researcher may be interested in interpreting them to explain how the synthetic control was built. For example, in the California tobacco control program application that we will re-visit in Section 5, the researcher may be interested in explaining what predictors (alcohol consumption, income etc) are used in generating the counterfactual. Hence, the choice of predictors is important both for the performance and the interpretability of the synthetic control.

Most of the literature relies on treating the predictor weights as a hyper-parameter and choosing them through a cross-validation procedure. Different implementations and best practices for choosing the weights have been proposed, see for example Abadie et al. (2010), Abadie et al. (2015), Doudchenko and Imbens (2016), and Ben-Michael et al. (2021). A common thread throughout, however, is that researchers should have a good prior on what predictors to include to construct the synthetic control. Recently, Pouliot and Xie (2022) note that cross-validation may not be amenable to many synthetic control settings and insightfully propose an information criteria for model selection of the donor weights based on estimating the degrees of freedom of the synthetic control problem. Under a gaussianity assumption they show that using the information criteria leads to the right selection of the combination of donor units.

In this paper our focus is slightly different. We are interested in the selection of the predictors themselves as well as how they relate to the performance of the synthetic control. Taking Abadie et al. (2015) as a starting point, we show theoretically and empirically that a modified penalized procedure that includes an l_1 penalty for the predictor weights can achieve consistent predictor selection as well as better mean squared error performance than the standard method. Our theoretical results for a linear factor model ensure that, almost surely, predictors that do not contribute to the outcome of interest in the model will not be used in the match to compute the synthetic control. Under stronger assumptions, the result also holds the other way around; predictors that are useful in generating the outcome of interest are used with positive probability. In the same setting we also derive a synthetic control bias bound akin to the one in Abadie et al. (2010) and Ferman and Pinto (2021), and show that the bias depends on whether the useful predictors can be matched well. Finally, we show that the proposed penalized procedure has a faster mean square error rate of convergence for matching the predictors, suggesting it may have lower bias than the standard method.

We refer to our proposed procedure as the *sparse* synthetic control as it is related to the class of penalized synthetic control models proposed by Quistorff et al. (2020). The focus of Quistorff et al. (2020), however, is on showing consistency and inference for penalized synthetic control estimators of the average treatment effect on the treated using the theoretical framework of Ferman (2021) and Chernozhukov et al. (2021). Our model consistency results also apply to the estimators proposed by Quistorff et al. (2020), therefore we see our theoretical results and proposed methodology as complementary to Quistorff et al. (2020) in the analysis of the properties of penalized synthetic control methods.

To evaluate the usefulness of the theoretical results in practice we compare in a simulation study the standard (un-penalized) synthetic control and our proposed method and revisit the passage of Proposition 99 in California using an extended data set with forty predictors. The simulation evidence confirms that the penalized method is able to identify the useful from the nuissance predictors and achieves better mean squared error in the post-treatment period. The empirical application highlights that when the number of predictors is large the standard un-penalized synthetic control may struggle to provide a valid synthetic control. On the other hand, our procedure is robust to increasing the number of predictors from the original seven used in Abadie et al. (2010) to a higher dimensional setting with forty predictors. Furthermore, the *sparse* synthetic control uses predominantly predictors included in the original application, confirming the predictor choice of Abadie et al. (2010). Finally, an estimate of the placebo variance in the application suggests that the proposed method has lower variance than the standard method, as predicted by the theory.

The paper is structured as follows. Section 2 describes the proposed penalized synthetic control method and a computation algorithm. Section 3 presents the main theoretical results in a linear factor model. Section 4 explains the simulation study results, and finally Section 5 discusses the empirical application to the California tobacco control program.

3.2 Sparse Synthetic Controls

To define the *sparse* synthetic control method, consider a setting in which we observe $J+1$ aggregate units for T periods. The outcome of interest is denoted by Y_{it} and only unit 1 is exposed to the intervention during periods $T_0+1,\ldots,T$. We are interested in estimating the treatment effect $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$ for $t > T_0$, where Y_{1t}^I and Y_{1t}^N denote the outcomes under the intervention and in absence of the intervention respectively. Since we do not observe Y_{1t}^N for $t > T_0$ we estimate τ_{1t} by building a counterfactual $\hat{Y}_{1t}^N$ of the treated unit's outcome in absence of the intervention.

As in the standard synthetic control our counterfactual outcome will be given by a weighted average of the donor units' outcomes, that is $\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}$ for a set of weights $\mathbf{w} = (w_2, \ldots, w_{J+1})'$. To choose the weight vector $\mathbf{w}$ we use observed characteristics of the units and pre-intervention measures of the outcome of interest. Formally, we let the $K \times 1$ design matrix for the treated unit be $\mathbf{X}_1 = (Z_1, \bar{Y}_1^{\mathbf{K}_1}, \ldots, \bar{Y}_1^{\mathbf{K}_M})'$, where $\{\bar{Y}_1^{\mathbf{K}_i}\}_1^M$ represent M linear combination of the outcome of interest for the pre-intervention period and Z_1 are pre-treament characteristics of the treated unit. Similarly, for the donor units, $\mathbf{X}_0$ is a $K \times J$ matrix constructed such that its jth column is given by $(Z_j, \bar{Y}_j^{\mathbf{K}_1}, \ldots, \bar{Y}_j^{\mathbf{K}_M})'$. We call the K rows of the design matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ the *predictors* of the outcome of interest. This can include, for example, lags of the outcome variable and important context dependent characteristics of the aggregate units averaged over the pre-treatment period.

We partition the pre-intervention period into a training set $(\mathbf{X}_0^{train}, \mathbf{X}_1^{train}, \mathbf{Y}_0^{train}, \mathbf{Y}_1^{train})$ for $t \in \{1, \ldots, T_v\}$ and a validation set $(\mathbf{X}_0^{val}, \mathbf{X}_1^{val}, \mathbf{Y}_0^{val}, \mathbf{Y}_1^{val})$ for $t \in \{T_v+1, \ldots, T_0\}$. This allows for the training and validation design matrices to differ, for example if predictors are averaged over different time periods or different linear combinations of lagged outcome variables are used.

The *sparse* synthetic control is defined by the tuple of weight vectors $(\mathbf{V}^*, \mathbf{w}^*)$ computed by solving the following bi-level optimization program:[1]

- Upper level problem:

$$(\mathbf{V}^*, \mathbf{w}^*) \in \operatorname{argmin}_{\mathbf{V},\mathbf{w}} L_V(\mathbf{V}, \mathbf{w}, \lambda) = \frac{1}{T_{val}} \|\mathbf{Y}_1^{val} - \mathbf{Y}_0^{val}\mathbf{w}(\mathbf{V})\|^2 + \lambda \|\mathrm{diag}(\mathbf{V})\|_1,$$
$$\text{s.t. } \mathbf{w}(\mathbf{V}) \in \psi(\mathbf{V}), \boldsymbol{V} \in \mathbb{R}_+^K.$$

- Lower level problem:

$$\psi(\mathbf{V}) \equiv \operatorname{argmin}_{\mathbf{w} \in \mathcal{W}} L_W(\mathbf{V}, \mathbf{w}) = \|\mathbf{X}_1^{train} - \mathbf{X}_0^{train}\mathbf{w}\|_V^2,$$

[1] I follow the notation of Malo, Eskelinen, Zhou and Kuosmanen 2020, who propose a computational method to solve a similar problem.

where,

$$\mathbf{w} \in \mathcal{W} \equiv \left\{\mathbf{w} \in \mathbb{R}^J \mid \mathbf{1}'\mathbf{w} = 1,\ w_j \geq 0,\ j = 2, \ldots, J+1\right\},$$
$$\mathbf{V} \in \mathcal{V} \equiv \left\{\mathbf{V} \mid \mathbf{V} \in \mathbb{R}^{K\times K},\ trace(\mathbf{V}) = 1,\ V_{kk} \geq 0, V_{kl} = 0 \text{ for } k \neq l\right\}, \text{ and}$$
$$\|\cdot\|_V \text{ denotes the semi-norm parameterized by } \mathbf{V} \text{ such that } \|\boldsymbol{A}\|_V = (\mathbf{A}'\mathbf{V}\mathbf{A})^{1/2}.$$

The main idea of the *sparse* synthetic control is that the l_1 penalty in the upper level problem induces some predictor weights (the diagonal elements of the weight matrix $\mathbf{V}$) to be set to zero as the penalty term λ increases. A similar bi-level program is also considered in Quistorff et al. (2020), in which the authors also consider additional penalties for the lower level problem. In practice, given the weight restriction $\mathbf{V} \in \mathcal{V}$, we follow Abadie et al. (2010) in the use of ex-post weight normalization by initially setting one predictor weight to one (i.e. $v_{k_0} = 1$ for some $k_0 \in \{1, \ldots, K\}$) and only restricting the v_k weights to be positive in the upper level program. In the appendix we describe why this ex-post normalization is necessary for finding a unique solution for the v weights. The following algorithm details the procedure used to choose the hyper-parameter λ and compute the *sparse* synthetic controls.

Result: $\mathbf{w}^*, \mathbf{V}^*$
Data: $(\mathbf{X}_0^{train}, \mathbf{X}_1^{train}, \mathbf{Y}_0^{train}, \mathbf{Y}_1^{train})$, $(\mathbf{X}_0^{train}, \mathbf{X}_1^{train}, \mathbf{Y}_0^{val}, \mathbf{Y}_1^{val})$
set $v_{k_0}= 1$;
initialize v_k for $k \neq k_0$ to $(\mathbf{X}_0^{train'}\mathbf{X}_0^{train})^{-1}$;
for *each λ in a grid* **do**
  get $(\mathbf{V}_\lambda, \mathbf{w}_\lambda)$ by jointly minimizing $L_W(\mathbf{V}, \mathbf{w}, \lambda)$ and $L_V(\mathbf{V}, \mathbf{w})$ for the training data;
    s.t. $\mathbf{w} \in \mathcal{W}$, $v_k \geq 0\ \forall k \neq k_0$ and $v_{k_0} = 1$;
  scale $\mathbf{V}_\lambda$ to $[0, 1]$;
  get $\mathbf{w}_\lambda^*$ by minimizing $L_W(\mathbf{V}_\lambda, \mathbf{w}, \lambda)$ for the training data;
  store $\text{MSE}(\mathbf{Y}_1^{val}, \mathbf{Y}_0^{val}\mathbf{w}_\lambda^*)$ and $\mathbf{V}_\lambda$;
end
choose λ^* with minimum $\text{MSE}(\mathbf{Y}_1^{val}, \mathbf{Y}_0^{val}\mathbf{w}_\lambda^*)$;
$\mathbf{V}^* = \mathbf{V}_{\lambda^*}$;
get $\mathbf{w}^*$ by minimizing $L_V(\mathbf{V}_\lambda^*, \mathbf{w})$ for the *shifted* training data.[a]

Algorithm 1: *Sparse* Synthetic Control

[a]The shifted training data is the training data but with time dependent variables shifted to the T_v periods before T_0.

3.3 Theoretical Results for a Linear Factor Model

To motivate the model selection procedure for the*sparse* synthetic controls theoretically, consider a standard setting in which the outcomes in absence of the intervention are given by a linear factor model as in Abadie et al. (2010)

$$Y_{it}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \epsilon_{it},$$

where δ_t is a common factor with equal loadings, Z_i is a $(k \times 1)$ vector of observed features, θ_t is a $(1 \times k)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, and ϵ_{it} is a unit-level transitory shock. Similar models have been used to motivate synthetic control and diff-in-diff estimators (see Arkhangelsky et al. (2021), Ferman and Pinto (2021) or Abadie and Vives-i-Bastida (2025)), as well as two-way fixed effect estimators and interactive fixed effect estimators (Bai (2009), Gobillon and Magnac (2016)). Assumption 1 imposes restrictions on the model primitives common in the literature.

Assumption 1 (Model primitives)**.** *Assumptions on the covariates, the factor structure and the error components.*

- *The covariates are fixed and bounded such that* $\max_j \|\boldsymbol{Z}_j\| \leq \sqrt{k}$.
- *The common factor* λ_t *are covariance stationary.*
- ϵ_{it} *are mean independent of* $\{\boldsymbol{Z}, \boldsymbol{\mu}\}$.

To study under which conditions the *sparse* synthetic control algorithm selects the most important predictors, I assume a sparse representation of the predictors.

Assumption 2 (Sparse representation)**.** *For all* t*,* $\boldsymbol{\theta}_t$ *is partitioned conformably into* $(\tilde{\boldsymbol{\theta}}_t, \mathbf{0})'$ *where* $\tilde{\boldsymbol{\theta}}_t$ *is a* $(k_1 \times 1)$ *vector of non-zero parameters and* $\mathbf{0}$ *is a* $k_2 \times 1$ *vector such that* $k = k_1 + k_2$.

Under Assumption 2 we partition the covariates comformably into a non-zero and zero group such that $\boldsymbol{Z}_i = (\boldsymbol{Z}_i^1, \boldsymbol{Z}_i^2)$. Throughout the paper I refer to the $\boldsymbol{Z}_i^1$ predictors as the "useful" predictors and the $\boldsymbol{Z}_i^2$ predictors as the nuisance predictors. Intuitively, we say that the *sparse* synthetic control is *interpretable* if given a large set of predictors the algorithm successfully recovers the useful ones. That is, we would like to require the method to consistently select which covariates are useful and give them positive weight, while giving zero weight to the nuisance covariates. Without further assumptions, however, this requirement is too strong. Indeed, if the nuissance covariates are highly correlated with the useful covariates then there is little hope to achieve model consistency. Assumption 3 gives a strong condition to avoid this problem.

Assumption 3 (Oracle covariate match)**.** *For fixed* J*, let the oracle weights be defined by*

$$\boldsymbol{w}^* \in argmin_{w \in \Delta^J} \mathbb{E}\|\mathbf{Y}_1 - \mathbf{Y}_0 \boldsymbol{w}\|^2.$$

We consider two assumptions:

1. *For all $k \in S = \{k \mid \theta_{tk} = 0$ for all $t\}$, $|Z_{1k} - Z'_{Jk}\boldsymbol{w}^*| > 0$.*

2. *(1) holds true and for $l \in S^c$, $|Z_{1l} - Z'_{Jl}\boldsymbol{w}^*| = 0$.*

Assumption 3 is different to the irrepresentability condition and the restricted eigenvalue condition common in compressed sensing and model selection consistency for penalized estimators (see Zhao and Yu (2006) and Chetverikov et al. (2016) for applications to lasso estimators). It says that the weights that minimize the statistical risk can not lead to a perfect match for a nuisance covariate. This assumption rules out data generating processes for which matching nuisance covariates improves the overall fit. For example, this could happen if the nuisance covariates are correlated with the factor loadings and contain useful information for matching them. Without this assumption, there could be cases in which a nuisance covariate is actually useful in improving the outcome fit and also can be matched at no cost to the lower level program, making it desirable for our algorithm to use it with positive probability.

The main result, Theorem 1, states that the *sparse* synthetic control algorithm is model consistent when the number of pre-treatment periods grows and J, k and F are fixed.

Theorem 1 (Model Selection)**.** *Under A1-A3.1 if ψ is an injective function and $\hat{\lambda} \to 0$ as $T_0 \to \infty$, for a fixed k and J, as $T_0 \to \infty$ the following holds*[2]

1. *If $k \in S = \{k \mid \theta_{tk} = 0$ for all $t\}$, then $P(v_k = 0) \to 1$.*

2. *If A3.2 holds and $l \in S^c$ then $P(v_l = 0) \to 0$.*

where v_k is the predictor weight for predictor m assigned by the sparse synthetic control algorithm.

In words Theorem 1 states that the *sparse* synthetic control algorithm will not use the nuisance predictors with probability going to one as the number of pre-treatment periods increases. The two main assumptions to derive the result are the oracle selection assumption and the assumption that the bi-level program has a unique optimum such that ψ is injective. This will be the case when the data points are in general position, that is, when every subset of the columns of $\boldsymbol{X}$ are linearly independent. These two assumptions are hard to test in practice.

It is important to note that the implication does not go the other way in general. The method could set a useful predictor weight to zero, but it will not assign a nuisance predictor a non-negative weight with high probability. Only if the useful predictors can be matched perfectly in the population, then the method is sign consistent. This result is important because it justifies the use of *sparse* synthetic controls to identify important predictors. A researcher that is unsure about what predictors to use to generate the synthetic control can use the *sparse* synthetic control with many predictors and be confident that it will use the "useful" predictors if the oracle separation assumption is satisfied.

[2]Here, $\hat{\lambda}$ is the cross-validated hyper-parameter of the penalty function denoted by λ^* in Algorithm 1.

This suggests using the method as an alternative to manual *predictor search* by researchers and possibly as a way to prevent specification search.

Next, we show that the *sparse* synthetic control also has desirable performance properties when compared to the standard (un-penalized) method. In particular, it has a faster convergence rate the for bias and mean-squared-error of the treatment effect estimator. To show this, recall that the treatment effect of interest is $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$ for $t > T_0$. To estimate it we generate a counterfactual $\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}^N$ for a synthetic control $\mathbf{w} = (w_2, \ldots, w_{J+1})'$. Therefore, the estimated treatment effect indexed by a synthetic control $\boldsymbol{w}$ is given by

$$\hat{\tau}_{1t}^{\boldsymbol{w}} = \boldsymbol{\theta}_t' \left(\boldsymbol{Z}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{Z}_j \right) + \boldsymbol{\lambda}_t' \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j \right) + \sum_{j=2}^{J+1} w_j (\epsilon_{1t} - \epsilon_{jt}).$$

It is well known that when the synthetic control does not have perfect pre-treatment fit it is biased. The following lemma gives an upper bound on the bias that depends on two terms that do not vanish asymptotically.

Lemma 1 (Bias Bound)**.** *Under A1-A2 and assuming that there exists a $\bar{\lambda}$ such that $|\lambda_{ft}| \leq \bar{\lambda}$ for all t and f, and that the smallest eigenvalue of $\sum_{t=1}^{T_0} \boldsymbol{\lambda}_t' \boldsymbol{\lambda}_t$ is bounded below by $\underline{\xi}$. Then, for a synthetic control $\boldsymbol{w}$,*

$$\mathbb{E}|\hat{\tau}_{1t}^{\boldsymbol{w}}| \leq \frac{\gamma}{T_0} \sum_{m=1}^{T_0} \mathbb{E}|Y_{1m} - \sum_{j=2}^{J+1} w_j Y_{jm}| + \left| \bar{\theta} \left(1 - \frac{\gamma}{T_0} \right) \right| \sum_{k=1}^{k_1} \mathbb{E}|Z_{1k}^1 - \sum_{j=2}^{J+1} w_j Z_{jk}^1| + O\left(T_0^{-1}\right).$$

where $\gamma = \left(\frac{\bar{\lambda}^2 F}{\underline{\xi}} \right)$, $\bar{\theta}$ is the maximum value of $\tilde{\boldsymbol{\theta}}_t$, F is the number of unobserved common factors and k_1 is the number of "useful" predictors.

The bias bound in Lemma 1 provides three insights. First, if we don't have perfect pre-treatment fit the bias does not vanish asymptotically as T_0 increases. This means that our treatment effect estimate will be biased even if we have many pre-treatment periods. Second, the bias depends on the mean absolute deviation of the pre-treatment outcomes. Pre-treatment fit is, therefore, very important for controlling the bias. This leads to the suggestion of not using synthetic controls when the pre-treatment fit is bad. Third, and most relevant for the *sparse* synthetic controls, the bias depends on the fit of the "useful" predictors $\boldsymbol{Z}_i^1$ and linearly in k_1. Therefore, even if we have a large number of pre-treatment periods and the pre-treatment fit is good (the first term in the bound is small), the bias could be large if the predictor fit is bad. Hence, a synthetic control that minimizes bias should attempt to perfectly match the useful predictors and disregard the nuisance predictors. This Lemma reinforces the result in Ferman and Pinto (2021) that synthetic controls with imperfect pre-treatment fit may in general be biased.

To formalize the improvement in terms of bias and mean-square-error of the *sparse* synthetic control over the standard synthetic control, we derive a finite sample rate for MSE($\boldsymbol{Z}_0 \boldsymbol{w}$) in the case in which the

Theorem 2 (MSE Rate)**.** *Let* $\boldsymbol{Z}_1 = \boldsymbol{Z}_0 w^* + \boldsymbol{u}$ *for* $u_i \sim_{ind} subG(\sigma_z^2)$*. Then, under* $A1 - A3.1$ *as* $T_0 \to \infty$*, almost surely for the sparse synthetic control* $\hat{\boldsymbol{w}}$*,*

$$MSE(\boldsymbol{Z}_1, \boldsymbol{Z}_0\hat{\boldsymbol{w}}) = \frac{1}{k}\|\boldsymbol{Z}_1 - \boldsymbol{Z}_0\hat{\boldsymbol{w}}\|^2 \lesssim \frac{\sigma_z\sqrt{k_1}}{k}\sqrt{2\log J}.$$

For the standard synthetic control $\tilde{\boldsymbol{w}}$*,*

$$MSE(\boldsymbol{Z}_1, \boldsymbol{Z}_0\tilde{\boldsymbol{w}}) = \frac{1}{k}\|\boldsymbol{Z}_1 - \boldsymbol{Z}_0\tilde{\boldsymbol{w}}\|^2 \lesssim \sigma_z\sqrt{\frac{2\log J}{k}}.$$

Theorem 2 describes how the mean squared error of the predictor match changes with the number of predictors and donor units. The main difference between the *sparse* synthetic control and the standard method is that the rate is of order $O(\sqrt{k_1}/k)$ instead of $O(1/\sqrt{k})$. This means that in a sparse setting, when k_1 is small with respect to k, the sparse synthetic control has a faster MSE rate. In practice, this is important because it implies that the sparse synthetic control will achieve lower MSE than the standard synthetic controls when both methods use the same number of predictors. More so, given Lemma 1, a faster predictor MSE rate also implies that the *sparse* synthetic control will achieve lower bias than the standard method. Formally, recall that by Cauchy-Schwartz inequality the MAD is bounded above by $\sqrt{J}MSE$, so the proposed method achieves a lower MAD, and so bias, than the standard method. In section 4 we explore these properties further by estimating the expected mean absolute deviations empirically and show in a simulation study that the *sparse* synthetic control achieves *both* better outcome pre-treatment fit and better "useful" predictor fit than the standard synthetic control.

3.4 Simulation Study

In this section we study the performance of the *sparse* synthetic control in relation to two benchmark models: the standard synthetic control with $\boldsymbol{V}$ chosen as $(\mathbf{X}_0^{train'}\mathbf{X}_0^{train})^{-1}$, which we label SCM, and the synthetic control with $\boldsymbol{V}$ chosen to minimize $\frac{1}{T_{val}}\|\mathbf{Y}_1^{val} - \mathbf{Y}_0^{val}\mathbf{w}(\mathbf{V})\|^2$ as proposed by Abadie et al. (2015), which we label SCM $\lambda = 0$ as it can be understood as the unpenalized version of the *sparse* synthetic control.

The findings below rely on the following simulation design, similar to the one used in Abadie and

Vives-i-Bastida (2025), for $B = 1000$ draws:[3]

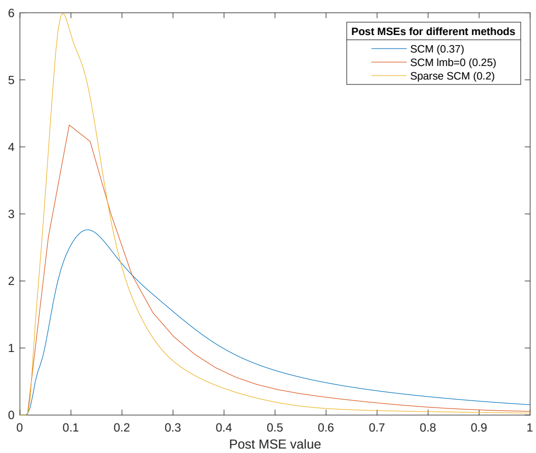
$$
\begin{aligned}
&T = 30,\ T_0 = 20,\ T_v = 10,\\
&\boldsymbol{Z}_i = [\boldsymbol{Z}_i^1,\ \boldsymbol{Z}_i^2],\ \text{where } \boldsymbol{Z}_i^1, \boldsymbol{Z}_i^2 \sim_{iid} U[0,1],\\
&\boldsymbol{Z}_1^1 = \frac{1}{2}\boldsymbol{Z}_2^1 + \frac{1}{2}\boldsymbol{Z}_3^1,\\
&\lambda_t \text{ follows an } AR(1) \text{ with coefficient } \rho = 0.5,\\
&\epsilon_{it} \sim N(0, \sigma^2) \text{ with } \sigma = 0.25,\\
&F = 7 \text{ in groups of 3 units and } J + 1 = 21.
\end{aligned}
$$

This simulation design implies that unit 1 (the treated unit) can be perfectly replicated (up to noise) by an average of units 2 and 3. Hence, the optimal synthetic control would choose $w_2 = w_3 = \frac{1}{2}$. We study two different predictor settings. The first is one in which there are a similar amount of useful and useless predictors ($k_1 = k_2 = 5$); the second, includes only one "useful" predictor ($k_1 = 1$ and $k_2 = 9$). In both cases we add 10 lags of the outcome variable to the design matrix, for a total of 20 predictors. Note that this is a challenging design for the method as both useful and useless predictors are drawn from the same distribution. We summarize the simulation results in two Figures that show the performance of the different methods and explore the theoretical insights from Section 3.

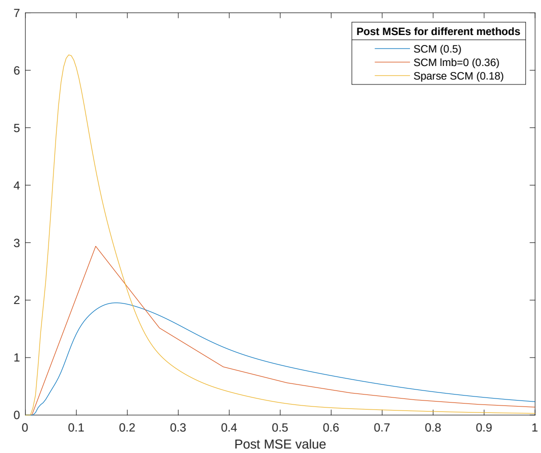
First, focus on the post-treatment mean squared error (MSE) of the outcome variable. Given that the MSE is an informative measure of fit that includes both bias and variance it gives us an idea of the performance of the estimator. In particular, lower post-treatment MSEs will imply lower standard errors for the treatment effect of interest. Figure 3.1 shows the distribution of MSEs for the simulations for the two settings. Panels (a) and (b) show that the *sparse* synthetic control has on average lower MSE and less dispersion than the two benchmark methods. Panels (c) and (d) show the distribution of the gaps between the predicted outcome and the real outcome for all periods. The *sparse* synthetic control has tighter confidence intervals both before and after treatment and sets unit weights closer to the optimal control (83% of the weight is given to units 2 and 3). Furthermore, observe that whereas the benchmarks methods perform worse in the $k_1 = 1, k_2 = 9$ setting, the *sparse* synthetic control is able to perform similarly in both settings. This shows the ability of the method to perform well regardless of the number of useful and nuisance predictors.

[3] We also set $\delta_t = 100$, but without loss of generality δ_t could be set to zero.

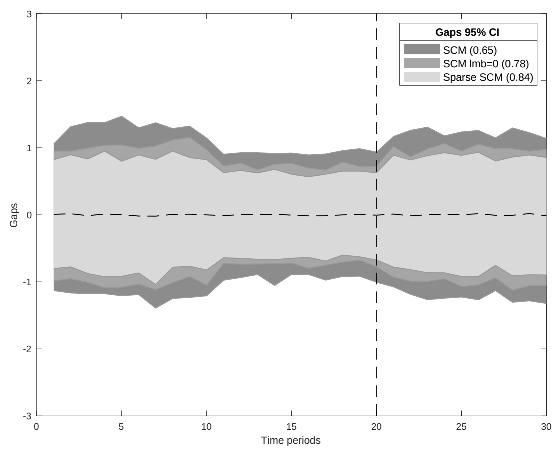
Figure 3.1: Mean Squared Errors.



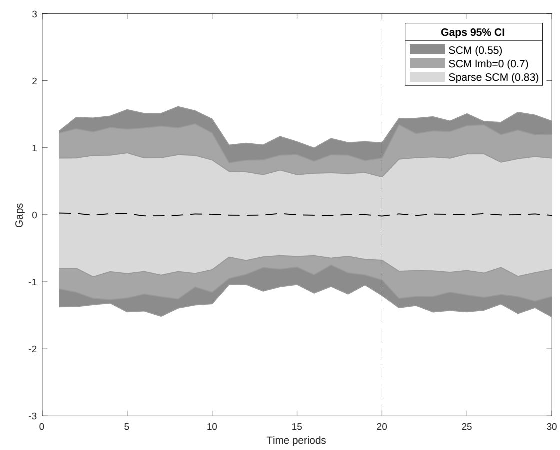
(a) $k_1 = k_2 = 5$.



(b) $k_1 = 1$, $k_2 = 9$.



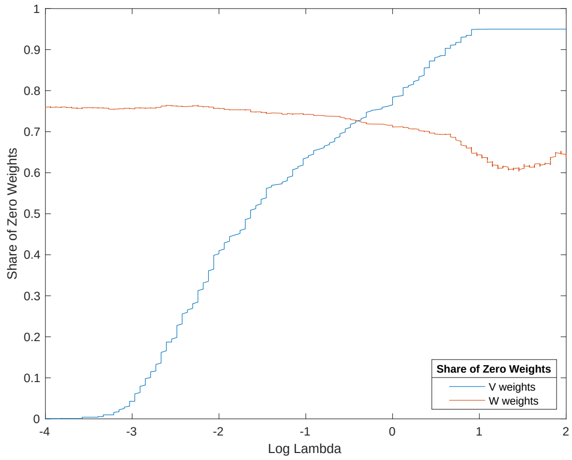
(c) $k_1 = k_2 = 5$.



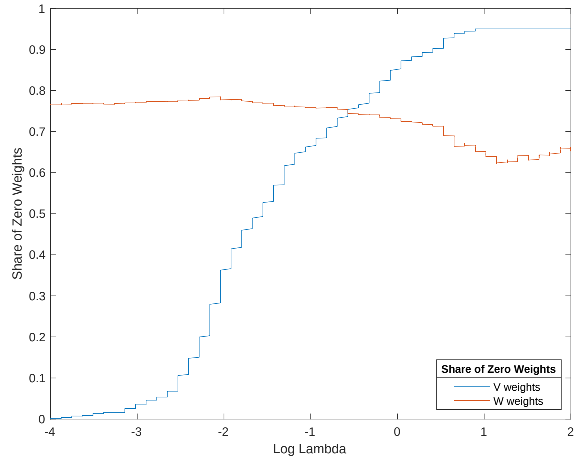
(d) $k_1 = 1$, $k_2 = 9$.

Notes: Panels (a) and (b) show the kernel density across simulations of the MSEs for the outcome variable in the post-treatment period, with average values in parenthesis. Panels (c) and (d) show the inter-quartile range of the $Y_{1t} - \hat{Y}_{1t}$ with $w_2^* + w_3^*$ in parenthesis. SCM lmb=0 refers to the unpenalized synthetic control with $\lambda = 0$.

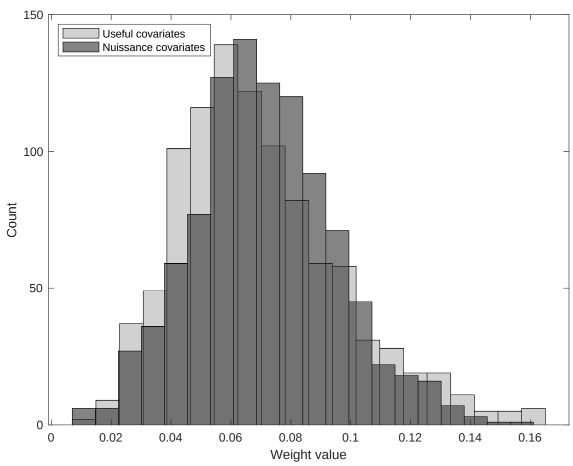
Recall that in section 3 the bias bound crucially depends on the MAD of the pre-treatment outcomes and useful predictors. A figure in the appendix shows the $\mathbb{E}$MADs for the outcome and useful predictors in the pre-treatment period. As in Figure 1, the *sparse* synthetic control is able to perform well in both settings whereas the benchmark methods perform poorly in the $k_1 = 1, k_2 = 9$ setting.



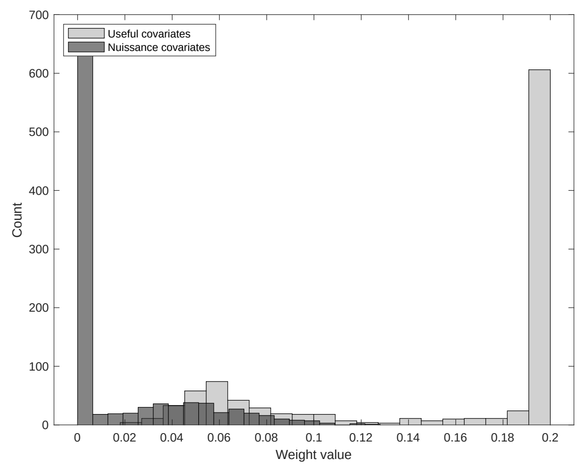
(c) $k_1 = k_2 = 5$.



(d) $k_1 = 1$, $k_2 = 9$.



(e) SCM $\lambda^* = 0$ $\boldsymbol{V}^*$



(f) *Sparse* SCM $\boldsymbol{V}^*$

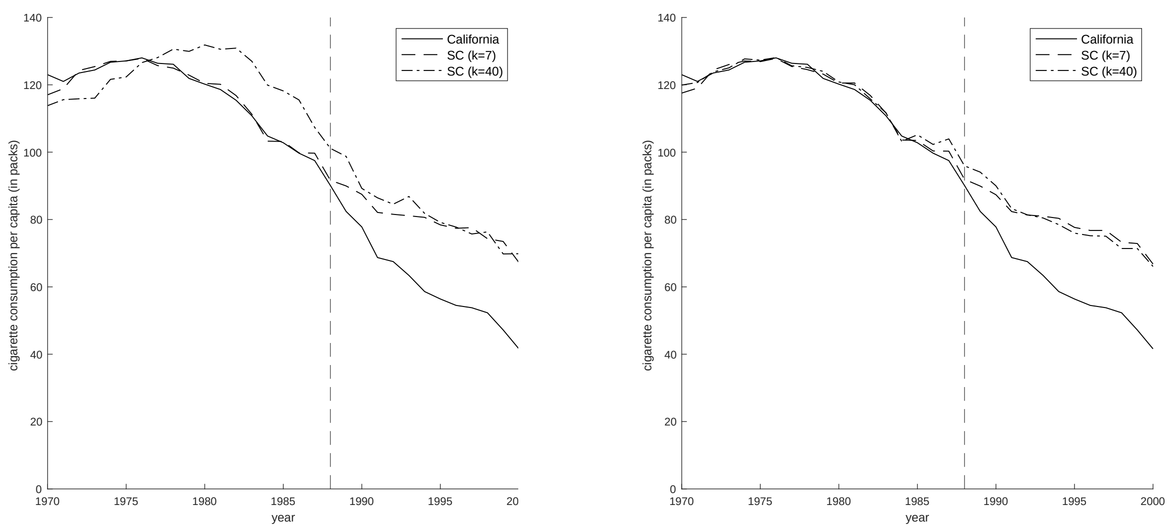
Notes: Panels (a) - (b) show the share of the $\boldsymbol{V}$ and $\boldsymbol{w}$ weights that are zero for the different values of λ^* across simulations for the *sparse* synthetic control. Panels (c) - (d) show the histogram of the v_k weights across simulations for the $k_1 = k_2 = 5$ setting for the unpenalized synthetic control with $\lambda = 0$ and the *sparse* synthetic control respectively.

To investigate the model selection result in Theorem 2 we plot the histogram of the predictor weight values (v_k) for the useful and useless predictors across simulations. Panels (c) and (d) in Figure 3.2 compare the SCM with $\lambda^* = 0$ and *sparse* synthetic control for the $k_1 = k_2 = 5$ setting. Whereas the standard SCM does not clearly distinguish between the useful and nuissance covariates, the *sparse* synthetic control correctly assigns zero weight to the nuissance covariates most of the time. The stark difference between the two models suggests that the *sparse* synthetic control may be able to perform variable selection in practice. Finally, in panels (a) and (b) in Figure 3.2 we show that the *penalized* synthetic control induces sparsity when the λ hyper-parameter increases, but that the donor weights (w_j) remain stable. The main takeaway is that the unit weights $\boldsymbol{w}^*$ have the same amount of sparsity regardless of the magnitude of the optimal regularization. This is evidence to motivate the technical assumption that ψ is injective in Theorem 2, and confirms that the method can reliably achieve a unique optimum. Note that we do see some instability for large values of λ^* in cases in which only one predictor is used.

3.5 Extending the California smoking program case study

In 1988 proposition 99 increased California's cigarette excise tax by 25 cents per pack and shifted public policy towards a clean air agenda. This policy intervention has been used extensively to compare the validity and performance of various synthetic controls and diff-in-diff estimators. The outcome of interest is cigarette sales per capita in packs in California and the donor pool includes 38 states without similar policy interventions. The original data set used in the Abadie et al. (2010) study used seven predictors: Ln(GDP per capita), percent aged 15–24, retail price, beer consumption per capita and three lags of cigarette sales per capita (1975, 1980 and 1988). In the study the standard synthetic control built using this design matrix falls outside the convex hull of the predictors, but has very good pre-treatment fit. This can be seen in Figure 3.3 panel (a) for the standard synthetic control with $k = 7$.

Figure 3.3: Synthetic Controls for California.



(a) Standard SC

(b) *Sparse* SC

To study the potential benefits of using penalized synthetic control methods, we augment the original predictor data set with 33 additional predictors. The additional predictors are obtained from the IPPSR (MSU) dataset assembled by Grossmann et al. (2021) of policy correlates in the United States, they include demographic variables, income related variables, political participation measures and government spending statistics. Importantly, they also include less useful variables such as state identifiers, but do not include predictors that do not vary across states as these would violate A3.1. Figure 3.3 compares the standard synthetic control (SCM with $\lambda^* = 0$) and the synthetic control calculated using Algorithm 1 using the $k = 7$ predictors considered in Abadie et al. (2010) and the extended set of predictors ($k = 40$).[4]

Figure 3.3 highlights the difficulty of the standard synthetic control matching problem when the number of predictors is large. While a good synthetic control exists in the $k = 40$ case, namely the same as the synthetic control for $k = 7$, the standard procedure may be unable to find a global optima and fail to recover the set of predictors for which a good match that concurrently optimizes pre-treatment fit exists. The synthetic control reported in Figure 3.3 panel (a) for $k = 40$ does not have good pre-treatment fit, and therefore, may lead to biased estimates of the average treatment effect on the treated. This example reinforces the notion in practice the choice of predictor set is very important in constructing valid synthetic controls. If researchers do not want to manually choose the predictors to match, or may not have a good prior on which predictors might be useful, the *sparse* synthetic control method we propose may be an attractive alternative. As can be seen in panel (b) of Figure 3.3 the sparse procedure is able to find a good synthetic control in the original and extended settings.

The theoretical results and simulation evidence suggest that the *sparse* synthetic control procedure should perform better in terms of bias and prediction mean squared error than the standard method in settings in which the predictors have a sparse representation. To explore this in the context of the California case study, Table 3.1 reports the average treatment effect on California of the passage of proposition 99 over the 11 post-treatment years and its placebo variance estimate for the Differences-in-Differences estimator using the rest of the US as control group, the standard synthetic control (SCM with $\lambda^* = 0$), and the *sparse* synthetic control for the original and extended predictor settings. The placebo variance is estimated using the placebo bootstrap method proposed in Arkhangelsky et al. (2021), a detailed explanation of the method can be found in the appendix.

To benchmark the findings in Table 3.1 in the first column we report the Differences-in-Differences estimate which, given that the parallel trends assumption is likely violated, is biased downwards as discussed in Abadie et al. (2010). The second column shows the standard synthetic control with the seven predictors as in Abadie et al. (2010), which given its excellent pre-treatment fit, is believed to have low bias. Column 3 shows that, unsurprisingly, the *sparse* method yields a very similar treatment effect as the standard method when $k = 7$. The difference between the methods become apparent in the last two columns. While the standard method becomes biased towards the differences-in-differences estimate when $k = 40$, the penalized method obtains a very similar estimate as in the $k = 7$ case. This robustness to the number of predictors highlights the benefits of the penalized procedures in choosing which predictors to match on to optimize the pre-treatment

[4]To choose the optimal predictor weights, we divide the pre-treatment period in a training period (14 years) and a validation period (5 years).

Table 3.1: Treatment effect estimates.

	DID	SCM	*Sparse* SCM	SCM	*Sparse* SCM
$\hat{\tau}$ estimate	-27.4	-18.9	-18.5	-21.0	-18.2
$\hat{V}_{\tau}^{1/2}$	(16.7)	(13.2)	(12.2)	(12.9)	(11.7)
k	-	7	7	40	40

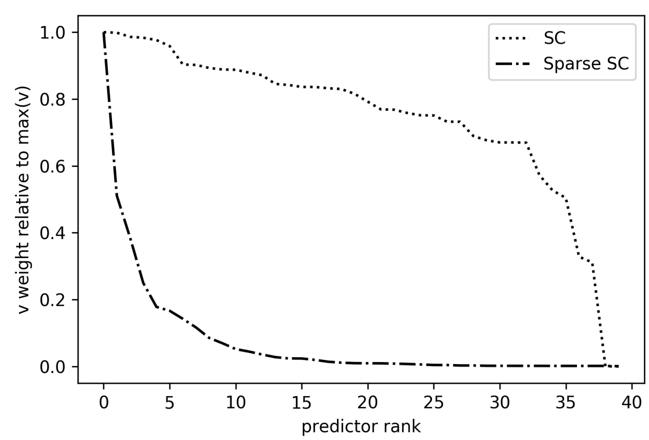
Notes: DID is the standard diff-in-diff estimator, SCM is the standard synthetic control with $\boldsymbol{V}$ chosen to minimize $\frac{1}{T_{val}}\|\mathbf{Y}_1^{val} - \mathbf{Y}_0^{val}\mathbf{w}(\mathbf{V})\|^2$ without penalization ($\lambda^* = 0$), SDID is the synthetic diff-in-diff estimator, and the $'+'$ indicates the augmented data setting. Standard errors are taken from Arkhangelsky et al. (2021).

Figure 3.4: Predictor choice in the $k = 40$ California case study.

(a) Top 7 predictors

SCM	*Sparse* SCM
smk_80	smk_75
general_revenue_inc	incshare_top1
smk_75	smk_88
smk_88	pc_inc_ann
loginc	region
general_expenditure_inc	budget_surplus
pc_inc_ann	taxes_gsp

(b) Predictor weight distribution



fit while avoiding over-fitting. Furthermore, we also see an improvement in the placebo variance estimates between the standard method and the penalized method, with the later having 8%-10% lower variance.

Finally, in Figure 3.4 we explore the predictor weight choices of each method when $k = 40$. Panel (b) shows that, as expected, the penalized method uses predominantly a small set of predictors, with a large fraction of predictors receiving zero or close to zero weight. This comes in stark opposition to the standard method that uses denser predictor weights, with many predictors having similar weight and only a few being set to zero. Panel (a) confirms that the choice of predictors in Abadie et al. (2010) was very good. Amongst the top seven predictors with highest weight, both methods use the lagged cigarette sales variables (smk) and personal income related variables log income and per capita annual income. Departing from the Abadie et al. (2010) predictor set both methods also assign weight to state level revenue and spending variables such as the general revenue, the budget surplus or the tax revenue as a percentage of the gross state product and both methods assign close to zero weight to the beer consumption measure. Additionally, the sparse method also uses a regional indicator (South-West-Midwest-East) and an inequality measure (the share of income by the top 1%). Finally, the sparse method assigns zero weight to the state indicator (a predictor we expected not to be useful in the match), while the standard method uses the indicator with positive weight. Suggesting that the model selection result for the sparse method is relevant in applied settings in identifying nuisance predictors.

The main takeaways from the empirical application are (1) that the standard SCM may struggle to generate a valid synthetic control, and so may be biased, when we use a large number of predictors, (2) that the *sparse* synthetic control is a potential solution to this problem and is robust to increasing the number of predictors, and (3) that using the penalized method in combination with a larger set of predictors may be useful for researchers that have many predictors at their disposal and are unsure about which ones may be the useful ones.

3.6 Conclusion

Researchers and policy makers are increasingly drawn towards synthetic control methods to analyze policy interventions. A key step in using these methods is deciding what predictors to use in building the synthetic control. Using a large number of predictors is not a valid alternative as it can lead to biased treatment effects and poor post-treatment performance. In this paper, we advocate for a data-driven penalized synthetic control method that automatically chooses the important predictors. We suggest that this method can be used as an alternative to the standard synthetic control when the researcher has many predictors at his disposal but does not know which ones should be used.

We show that, in a linear factor model setting, that the *sparse* synthetic control is model consistent and can successfully recover which predictors are useless when the number of pre-treatment periods is large. Furthermore, by deriving an MSE convergence rate result and a bias bound result, we show that the *sparse* synthetic control has better theoretical performance properties than the un-penalized method. Motivated by this insight we then show through a simulation study that the proposed method is able to reduce both the bias and the mean-squared error measures with respect to the un-penalized synthetic control.

Finally, we highlight the practical relevance of the method by applying it to the passage of Proposition 99 in California in a setting with 8 predictors versus a setting with 40 predictors. Whereas the standard synthetic control estimate becomes more biased when the number of predictors is increased, the *sparse* synthetic control is robust to the number of predictors used. Through this exercise, we confirm that the choice of predictors in the original study of Abadie et al. (2010) was good, as the penalized method uses the original predictors with positive probability.

A natural next step for future work is to explore alternative penalized synthetic control methods with multidimensional shrinkage (Vives-i-Bastida (2021)), for example with elastic net penalties, that may further improve the performance of the method and the standard errors of the treatment effect estimates.

Chapter 4

Synthetic Controls in Action

written jointly with Alberto Abadie

4.1 Introduction

Synthetic control methods have become widely applied in empirical research in economics and other disciplines. Recent empirical applications of the synthetic control framework include studies of the effects of political connections (Acemoglu et al., 2016), legalized prostitution (Cunningham and Shah, 2018), right-to-carry laws (Donohue et al., 2019), and many other (see Abadie, 2021, for a more elaborate review of the applications of synthetic control estimators). With the increasing popularity of synthetic control estimators, it has become particularly important to understand the settings where synthetic controls produce reliable estimates and those where they do not.

In this article, we explore in detail how the properties of synthetic controls translate to actual performance. From the properties of synthetic control estimators we distill a set of simple principles to guide empirical practice, and demonstrate the practical relevance of our proposed principles via simulation exercises.

First, we use a variety of data configurations to demonstrate the crucial role of pre-treatment fit in the performance of synthetic control estimators. Close pre-treatment fit, however, does not guarantee good performance of synthetic control estimators because of the possibility of over-fitting. We describe the nature of over-fitting within the context of synthetic control estimation, explain how over-fitting induces biases, demonstrate the practical relevance of these biases, and discuss how a careful study design can minimize or avoid them. Finally, we discuss validation exercises that can be applied to assess the validity of synthetic control estimators in empirical applications.

The rest of the article is organized as follows. Section 4.2 introduces synthetic control estimators and discusses their formal properties under a linear factor structure in the data generating process. Section 4.3 employs a grouped version of the linear factor model of Section 4.3 to study how the performance of synthetic controls is related to pre-treatment fit and over-fitting. Section 4.4 discusses how to validate synthetic control estimates. Section 4.5 discusses trimming as a way to alleviate over-fitting and interpolation biases. Section 4.6 discusses the role of covariates. The applicability of synthetic control estimators is not confined, however, to the linear factor model. In Section 4.7, we use a simple auto-regressive model to illustrate this point. Section 4.8 concludes.

4.2 A Primer on Synthetic Controls

Suppose we observe $j = 1, \ldots, J+1$ aggregate units, such as states or countries for $t = 1, \ldots, T$ periods. The first unit (that is, $j = 1$) is exposed to a policy intervention, or some other event or treatment of interest, at time $t = T_0 + 1$, with $T_0 + 1 \leq T$. The remaining J units are not exposed to the event or intervention of interest. We aim to estimate the effect of the treatment on some outcome of interest during the post-treatment periods, $T_0 + 1, \ldots, T$. We will use the terms "intervention", "event", and "treatment" interchangeably.

To define treatment effects, we formally adopt a model of potential outcomes (Rubin, 1974).

Let Y_{jt}^N be the potential outcome observed for unit $j \in \{1, \ldots, J+1\}$ and time $t = \{1, \ldots, T\}$ in the absence of the intervention. Let Y_{1t}^I be the potential outcome observed for the treated unit at time $t = T_0 + 1, \ldots, T$ under the intervention. For each unit and time period, Y_{jt} is the observed outcome. Therefore, observed outcomes for untreated units, $j = 2, \ldots, , J+1$, are equal to Y_{jt}^N. For the treated unit, the observed outcome is equal to Y_{1t}^N for $t = 1, \ldots, T_0$, and equal to Y_{1t}^I for $t = T_0 + 1, \ldots, T$. The object of interest is the treatment effect on the treated unit,

$$\tau_t = Y_{1t}^I - Y_{1t}^N,$$

for $t = T_0 + 1, \ldots, T$. Because $Y_{1t}^I = Y_{1t}$ in the post-treatment periods, we obtain

$$\tau_t = Y_{1t} - Y_{1t}^N,$$

for $t = T_0 + 1, \ldots, T$. That is, because Y_{1t}^I is observed in the post-treatment periods, estimating τ_{1t} for $t = T_0 + 1, \ldots, T$ boils down to estimating Y_{1t}^N. A synthetic control estimator of Y_{1t}^N is a weighted average of the outcomes for the "donor pool" of J untreated units,

$$\widehat{Y}_{1t}^N = \sum_{j=2}^{J+1} W_j Y_{jt},$$

where $W_2, \ldots, W_{J+1}$ are non-negative and sum to one. A synthetic control estimator of τ_{1t} is equal to the difference between the outcome values for the treated units and the outcomes values for the synthetic control,

$$\widehat{\tau}_t = Y_{1t} - \sum_{j=2}^{J+1} W_j Y_{jt}.$$

The weights, $W_2, \ldots, W_{J+1}$, represent the contribution of each untreated observation to the estimate of the counterfactual of interest, $\widehat{Y}_{1t}^N$. While these weights could, in principle, be directly chosen by the analyst, the synthetic control literature provides a host of data-driven weight selectors. For $j = 1, \ldots, J+1$, let $X_j = (X_{1j}, \ldots, X_{kj})'$ be a $(k \times 1)$-vector of pre-intervention values of predictors of Y_{jt}^N, with $t = T_0 + 1, \ldots, T$. Let X_0 be the $(k \times J)$-matrix that concatenates $X_2, \ldots, X_{J+1}$. A simple data-driven selector $W^* = (W_2^*, \ldots, W_{J+1}^*)'$ of the synthetic control weights minimizes

$$\|\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}\| = \left(\sum_{h=1}^{k} v_h (X_{h1} - W_2 X_{h2} - \cdots - W_{J+1} X_{hJ+1})^2 \right)^{1/2}, \tag{4.1}$$

subject to the constraints that the weights are non-negative and sum up to one. The non-negative constants, $v_1, \ldots, v_k$, can be used to standardize the predictors (for example, by making v_h equal to the inverse of the variance of the h-th predictor) or chosen to reflect their predictive power on the outcome of interest.

Synthetic control estimates, as defined above, are typically sparse, meaning that only a few units in the donor pool obtain weigths, W_j^*, different from zero. Figure 4.1 provides a geometric interpretation of this property. The synthetic control $\boldsymbol{X}_0\boldsymbol{W}^*$ is given by the projection of $\boldsymbol{X}_1$ on the convex hull of the columns of $\boldsymbol{X}_0$. The curse of dimensionality implies that, even in moderately large dimension, $\boldsymbol{X}_1$ may often be outside the convex hull of the columns of $\boldsymbol{X}_0$. This is the case depicted in Figure 4.1. As indicated in the figure, when $\boldsymbol{X}_1$ is outside the convex hull of the columns of $\boldsymbol{X}_0$, the synthetic control $\boldsymbol{X}_0\boldsymbol{W}^*$ is unique and sparse provided that the columns of $\boldsymbol{X}_0$ are in general position (see Abadie and L'Hour, 2021, for the meaning of "general position"). Then, the number of non-zero weights, W_j^*, is not larger than k, the dimension of $\boldsymbol{X}_j$. In Figure 4.1, the only untreated units that contribute to the synthetic control are represented by the three vertices of the shaded facet, the one that contains $\boldsymbol{X}_0\boldsymbol{W}^*$. If $\boldsymbol{X}_1$ is in the convex hull of the columns of $\boldsymbol{X}_0$, then the synthetic control does not need to be unique or sparse. However, sparse solutions with no more than $k+1$ non-zero weights exist by Carathéodory's theorem. Abadie and L'Hour (2021) provide a penalized synthetic control estimator that is unique and sparse even in the case that $\boldsymbol{X}_1$ is in the convex hull of the columns of $\boldsymbol{X}_0$. The weight selector in Abadie and L'Hour (2021) favors synthetic controls composed by untreated units with values of the predictors in $\boldsymbol{X}_j$ close to the values of the predictors for the treated unit, $\boldsymbol{X}_1$. The result is a procedure that not only produces unique and sparse estimates, but also ameliorates potential interpolation biases that could emerge from averaging outcomes between untreated units that are far from the treated unit in the space of the predictors, $\boldsymbol{X}_j$. Sparsity plays an important role in facilitating the interpretability of synthetic control estimates. The exact nature of the synthetic control estimate, the contribution of each unit in the donor pool to this estimate, and the size and direction of potential biases that may arise if the untreated units contributing to a synthetic control are indirectly affected by the treatment or by other known idiosyncratic shocks to their outcomes, are greatly facilitated by the non-negativity and sparsity of the weights and by the fact that they sum up to one, so they can be interpreted as proper weights.

To study the properties of synthetic control estimators, we will posit a generative model for the outcomes, Y_{jt}^N. Abadie et al. (2010) consider the linear factor model,

$$Y_{jt}^N = \delta_t + \boldsymbol{\theta}_t\boldsymbol{Z}_j + \boldsymbol{\lambda}_t\boldsymbol{\mu}_j + \epsilon_{jt}, \tag{4.2}$$

where δ_t is a time trend, and Z_j and μ_j are vectors of observed and unobserved predictors, respectively, with time varying coefficients, θ_t and λ_t. We will also use the term "covariates" to refer to Z_j and μ_j. The variable ϵ_{jt} is a transitory shock that we will model as white noise. We will take all the components on the right-hand side of equation (4.2) as given (conditioned on) except for ϵ_{jt}.

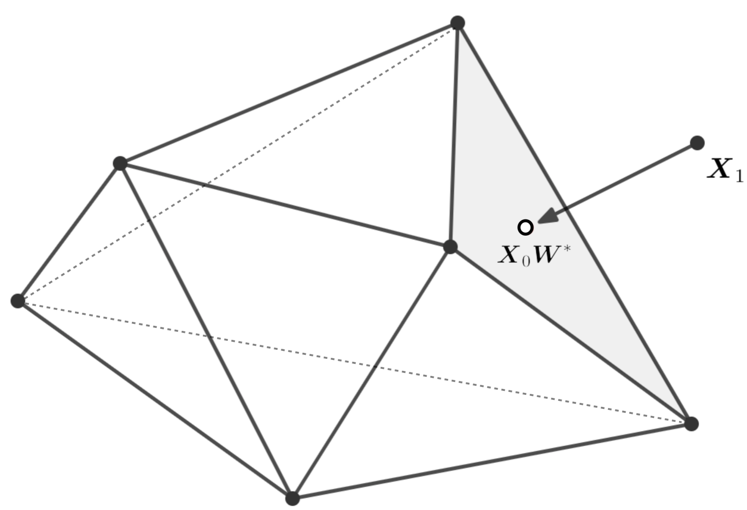
Equation (4.2) imposes a zero mean restriction on ϵ_{jt} regardless of treatment status, which can be interpreted as a form of unconfoundedness conditional on Z_j and μ_j. However, equation (4.2) does not directly restrict the process that determines selection for treatment. In the contexts where synthetic controls are typically applied in practice—with only one or a small number of treated units—a precise knowledge of the selection mechanism may not be particularly useful in the absence of a model for the outcomes. Consider the usual setting with only one treated unit. Randomization of the treatment–something that is extremely unlikely in settings with aggregate units, like states or countries—implies that a simple difference in means between outcomes of treated and untreated units is ex-ante unbiased for the average treatment effect (that is, unbiased before randomization). However, a simple difference in means could be heavily biased ex-post if randomization produces large differences between the characteristics of the treated unit and the units that are not exposed to the treatment. Equation (4.2) accounts for these differences by modeling the effect of observables, Z_j, and unobservables, μ_j, on the potential outcomes, Y_{jt}^N.

Suppose that, with probability one, the solution to the minimization of (4.1) subject to the weight constraints yields

$$\sum_{j=2}^{J} W_j^* Z_j = Z_1 \quad \text{and} \quad \sum_{j=2}^{J} W_j^* Y_{jt} = Y_{1t}, \tag{4.3}$$

for $t = 1, \ldots, T_0$. Under these conditions, a bound can be established on the magnitude of the bias, $|E[\widehat{\tau}_t - \tau_t]|$, of the synthetic control estimator. Abadie et al. (2010) provide a precise expression for this bound. For the purpose of the present article, however, it is enough to know that the bound increases with *(i)* the ratio between the scale of the transitory shocks, ϵ_{jt}, and the number of pre-intervention periods, T_0, *(ii)* the number of units in the donor pool, J, and *(iii)* the dimension of μ_j (i.e., the number of unobserved factors). The bias bound and equation (4.3) motivate including Z_j and pre-treatment values of the outcomes in the vector of matching variables, X_j. This equation may only hold approximately in practice if X_1 is outside the convex hull of the columns of X_0.

Figure 4.1: Sparsity of Synthetic Controls: A Geometric Interpretation.



Note: Projection of $\boldsymbol{X}_1$ on the convex hull of the columns of X_0.

We next provide intuition for why *(i)* to *(iii)* may affect the size of the bias. Notice that, under the model in (4.2), synthetic control weights that reproduce the values of observed and unobserved covariates,

$$\sum_{j=2}^{J} W_j^* Z_j = Z_1 \quad \text{and} \quad \sum_{j=2}^{J} W_j^* \mu_j = \mu_1, \tag{4.4}$$

would yield an unbiased estimator of the treatment effect. Although the unobserved covariates, μ_j, cannot be fitted directly, a synthetic control such that (4.3) holds employs pre-treatment outcomes as proxies for the unobserved factors. This would clearly be justified if the scale of the transitory shocks is small, so most of the heterogeneity in pre-treatment outcomes that does not come from Z_j is generated by heterogeneity in μ_j. In that case, fitting pre-treatment outcomes, as in (4.3), comes very close to fitting the unobserved factors, as in (4.4). However, a substantial amount of variability in ϵ_{jt} opens the door to over-fitting: the possibility that pre-treatment outcomes are fitted in (4.3) out of variation in ϵ_{jt}. In that case, the fitted value for the unobserved factors given by the synthetic control,

$$\sum_{j=2}^{J+1} W_j^* \mu_j,$$

may substantially differ from μ_1, inducing bias. The probability that (4.3), or an approximate version of it, is mainly produced by over-fitting is small when the scale of ϵ_{jt} is small and the number of pre-treatment periods, T_0, to be fitted is large. A large donor pool, however, increases the chances that the pre-treatment outcomes are fitted out of variation in ϵ_{jt}. Moreover, given that the linear model in (4.2) is nothing but a local approximation to a data generating process that could be non-linear, fitting the values of the treated units with untreated units that are far from each other in the space of the predictors may result in sizable interpolation biases. Finally, the dimension of μ_j reflects the variability induced by unobserved covariates. In the absence of unobserved covariates, a synthetic control that reproduces the value of Z_1 only (as in the first part of (4.4)) would be exactly unbiased.

From our discussion of the synthetic control design and the bias bound, we can now distill seven guiding principles for empirical practice with synthetic control estimators:

1. Low volatility. Closely fitting a highly volatile series is likely the result of over-fitting at work, especially if it happens over a short pre-intervention period. Synthetic controls were designed for settings with aggregate series, where aggregation attenuates the magnitude of the noise.

2. Extended pre-intervention period. A good control unit must closely reproduce the trajectory of the outcome variable for the treated unit over an extended pre-intervention period. A good fit, if it is the result of a secular agreement between the treated and the synthetic control units in their responses to unobserved factors, should persist in time.

3. Small donor pool. A larger donor pool is not necessarily better than a smaller one. Adopting a small donor pool of untreated units that are close to the treated unit in the space of the predictors helps reduce over-fitting and interpolation biases.

4. Sparsity makes synthetic controls interpretable.

5. Covariates matter. A component of Z_i that is not controlled for (that is, not included in X_j) is effectively thrown into μ_j. Fitting Z_j is easier than fitting μ_j.

6. Fit matters. The bound on the bias is predicated on close fit. A deficient fit raises concerns about the validity of a synthetic control (see, however, Ferman, 2021; Ferman and Pinto, 2021, for exceptions and qualifications on this rule).

7. Out-of-sample validation is key. The goal of synthetic controls is to predict the trajectory of the outcome variable for the treated unit in the absence of the intervention of interest. The quality of a synthetic control can be assessed by measuring predictive power in pre-intervention periods that are left out of the sample used to calculate the synthetic control weights.

We will study and illustrate the empirical relevance of the seven principles described above. Principles 1-6 derive directly from the bias bound in Abadie et al. (2010). Because the formal validity of the bias bound rests on strong assumptions, it is useful to investigate the extent to which these principles effectively translate to empirical practice. We will also illustrate the effectiveness of validation techniques derived from the last principle on the list.

We will employ simulations under a variety of data generating processes to investigate how features of the data affect the performance of synthetic control estimators in ways that can be predicted from the seven guiding principles above. An important take-away of our analysis is that, as previously argued in Abadie (2021) and others, contextual and data requirements are key for the performance of synthetic control estimators, and should be carefully checked in empirical applications. Mechanistic applications of the method, without regard to the guiding principles described above, leave the results of the empirical exercise vulnerable to biases created by over-fitting and by discrepancies between the values of the predictors for the treated units and for the units that contribute to the synthetic control. When conscientiously applied, however, synthetic controls become powerful and transparent design tools for estimating the effects of aggregate interventions.

4.3 Performance of the Synthetic Control Estimator

To begin our investigation of the performance of the synthetic control estimator under a variety of features in the data, we adopt the grouped factor model of Ferman (2021), which is a special case of (4.2). We will use a generative model with F equally-sized groups of units denoted $f(i) \in \{1, \ldots, F\}$ and F common factors, with each unit i loading exclusively on factor $f(i)$,

$$Y_{it}^N = \delta_t + \lambda_{f(i)t} + \epsilon_{it}. \tag{4.5}$$

The common factors, λ_{ft}, follow an $AR(1)$ processes with autoregressive coefficient ρ and standard Gaussian innovations. ϵ_{it} follows a Gaussian distribution, with mean zero and variance σ^2, independent of any other component.

We assume that only the first unit is treated and that only the second unit loads on the same factor as the treated unit, so $f(1) = f(2)$. The other untreated units—that is, units $j = 3, \ldots, J+1$—are also grouped into pairs that load on a pair-specific factor, $f(3) = f(4)$, $f(5) = f(6)$, and so on. An unbiased synthetic control in this setting is one with $W_2^* = 1$ and $W_3^* = W_4^* = \cdots = W_{J+1}^* = 0$. This paired design represents the least favorable case for synthetic control estimators under the grouped factor model of Ferman (2021): there is only one unit that reproduces the trajectory of $\lambda_{f(i)t}$ for the treated unit in the absence of the intervention.

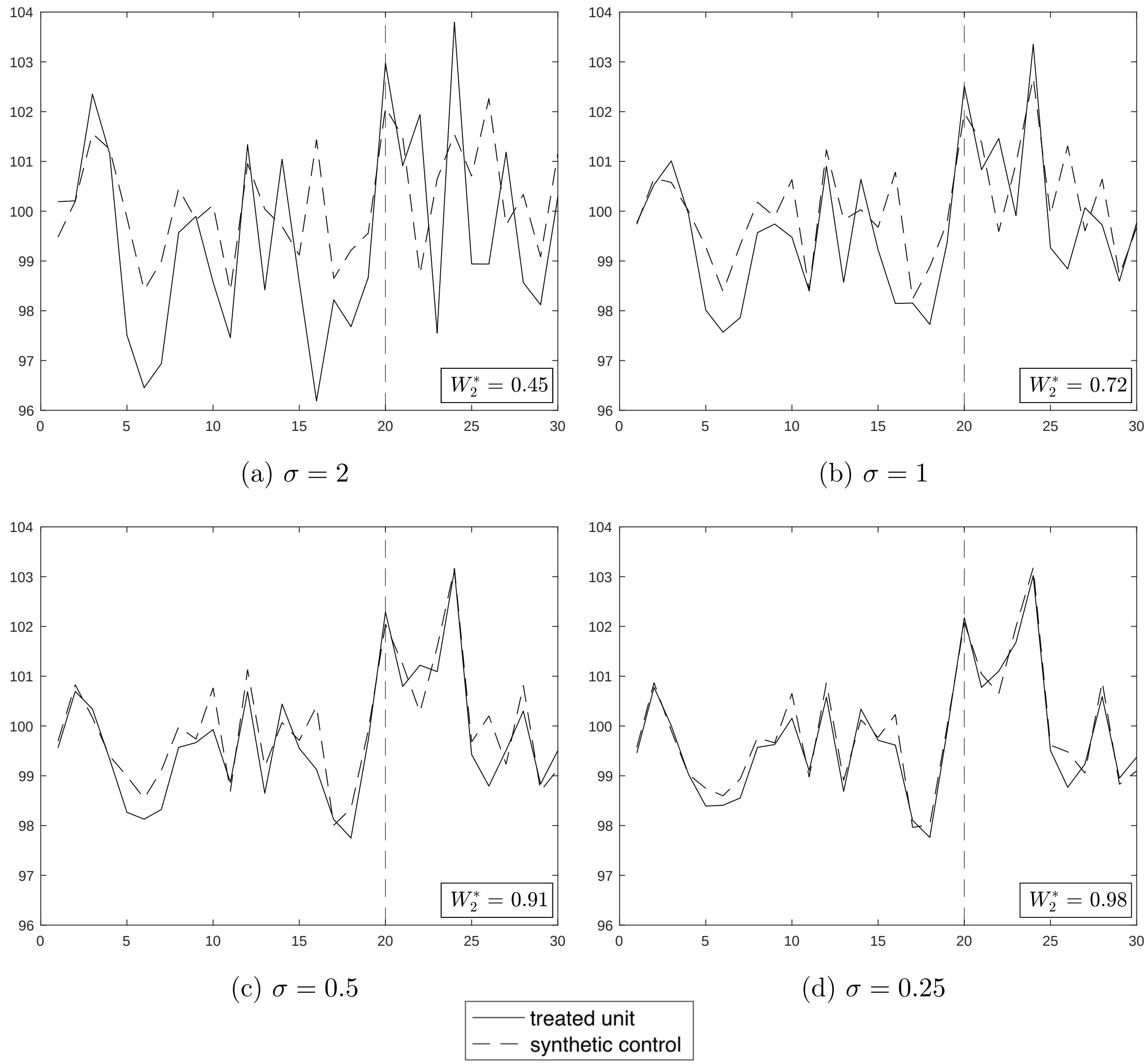
Unless we state otherwise, we will impose $Y_{1t}^I = Y_{1t}^N$ for $t = T_0 + 1, \ldots, T$, so the treatment effect on the treated unit is equal to zero. This allows us to interpret deviations from zero in treatment effect estimates as reflective of lack of accuracy or precision. We fix $T = 30$ and study the performance of the synthetic control estimator (and the extent to which the estimated weights approximate $W_2^* = 1$) under a variety of values for σ, T_0, ρ and F. We will also investigate the extent to which variations on the basic estimator improve performance or diminish it.

Figure 4.2 reports simulations of the synthetic control estimator for the case $T_0 = 20$, $\rho = 0.5$, and $J = 19$ (so $F = 10$). Each panel contains results for a different value for σ. With a large value of σ in panel (a), the pre-treatment fit is poor, which translates in large estimation errors during the post-treatment periods. As σ decreases in the subsequent panels, the pre-treatment fit improves noticeably and it translates into similar precision gains during the post-intervention periods. These gains in estimation accuracy are reflected in substantial increases in the value of W_2^* as σ decreases. With $\sigma = 2$ in panel (a), the synthetic control assigns less than half of the weight the the "correct" unit, $j = 2$. When $\sigma = 0.25$ in panel (d), the weight of unit $j = 2$ on the synthetic control estimator is almost one. The pattern of results in Figure 4.2 is reflective of the fact that the bias of the synthetic control estimator and the quality of the pre-treatment fit both depend on the scale of the individual transitory shocks, ϵ_{jt}. Under the generative process in (4.2), lack of pre-treatment fit can arise because of noise in the series (large σ) or because the values of Z_1 and μ_1 cannot be closely reproduced with a convex combination of the values of Z_j and μ_j for the units in the donor pool. In both cases, post-treatment synthetic control estimates could incorporate sizable biases. The results in Figure 4.2 underscore the importance of a good pre-treatment fit.

In order to visually demonstrate the performance of synthetic controls "in action", Figure 4.2 reports the results for one random realization in our simulation design only. One could, instead, run a large number of simulations and plot or tabulate results on pre-treatment fit and post-treatment bias. Figure 4.3 plots results for 10000 simulations, with the same simulation design as in Figure 4.2. Like Figure 4.2, the new Figure 4.3 indicates substantial increases in estimation accuracy that are associated with pre-intervention fit. For the rest of this article, we report results of single random simulations in the main text of the article. In the appendix we report the analogous results calculated over a large number of simulations.

Figure 4.4 shows the results of a simulation that employs the same data generating process as in

Figure 4.2: Pre-treatment fit and estimation error.

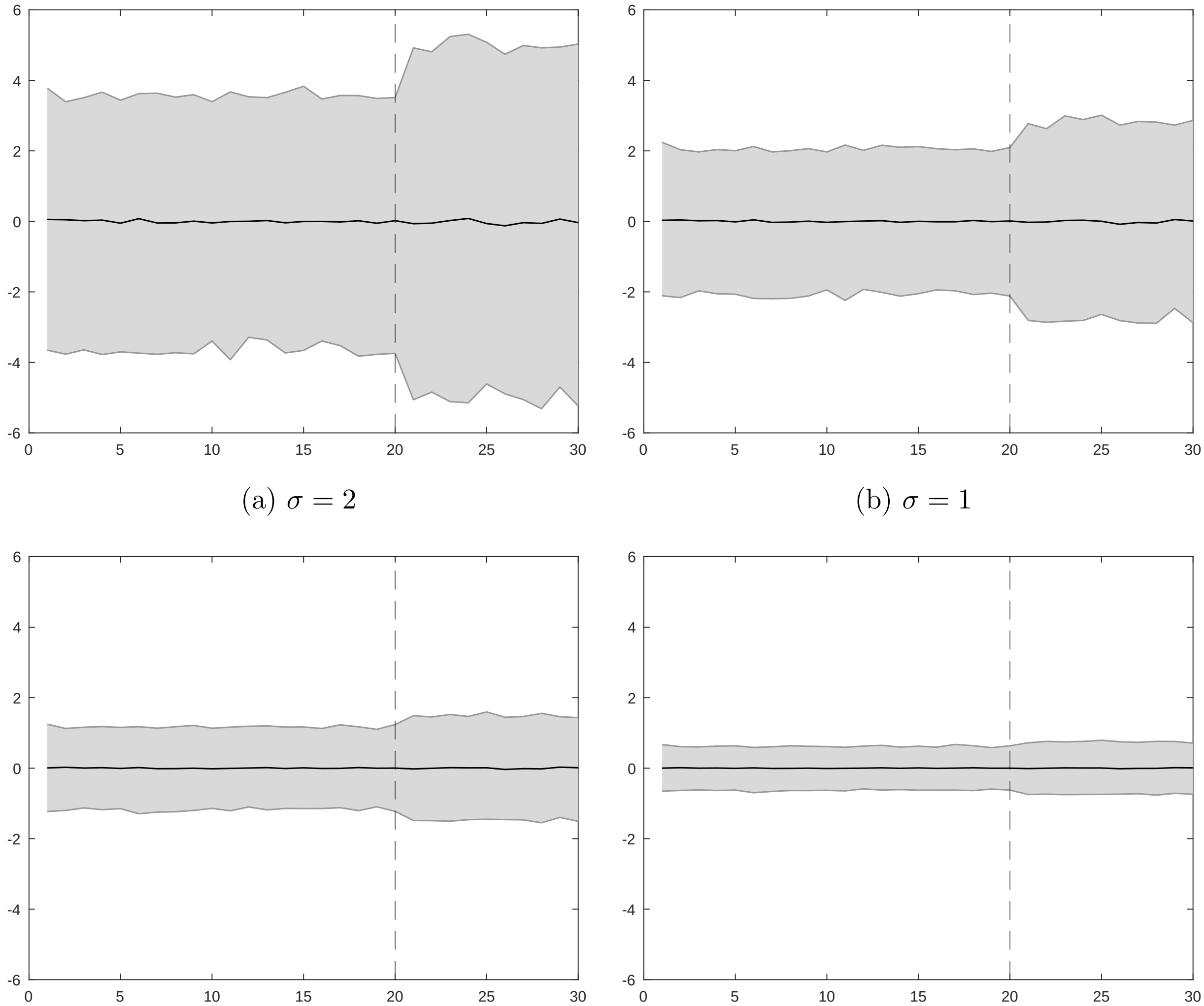


(a) $\sigma = 2$ (b) $\sigma = 1$ (c) $\sigma = 0.5$ (d) $\sigma = 0.25$

Note: Synthetic control estimates for the grouped factor model in equation (4.2), with $\rho = 0.5$, $T_0 = 20$, and $J + 1 = 20$ ($F = 10$). Each panel reports results for a different value of σ.

Figure 4.2 with the exception that now the series λ_{ft} incorporate stochastic trends. As for Figure 4.2, the magnitude of the noise in ϵ_{jt} is key for pre-treatment fit and post-treatment prediction error. Notice, however, that heterogeneity in the trending behavior of the series helps with the selection of a synthetic control heavily based on the outcome series of unit 2, which is affected by the same stochastic trend as the treated unit. In addition to the outcome series for the treated unit and the synthetic controls, Figure 4.4 also reports the average outcome path for all untreated units. Figure 4.4 shows large gains from synthetic control estimation, relative to a simple average of the outcome for the untreated units. These gains are created by heterogeneity across units in the outcome trends,

Figure 4.3: Prediction error over 10000 simulations



(a) $\sigma = 2$

(b) $\sigma = 1$

(c) $\sigma = 0.5$

(d) $\sigma = 0.25$

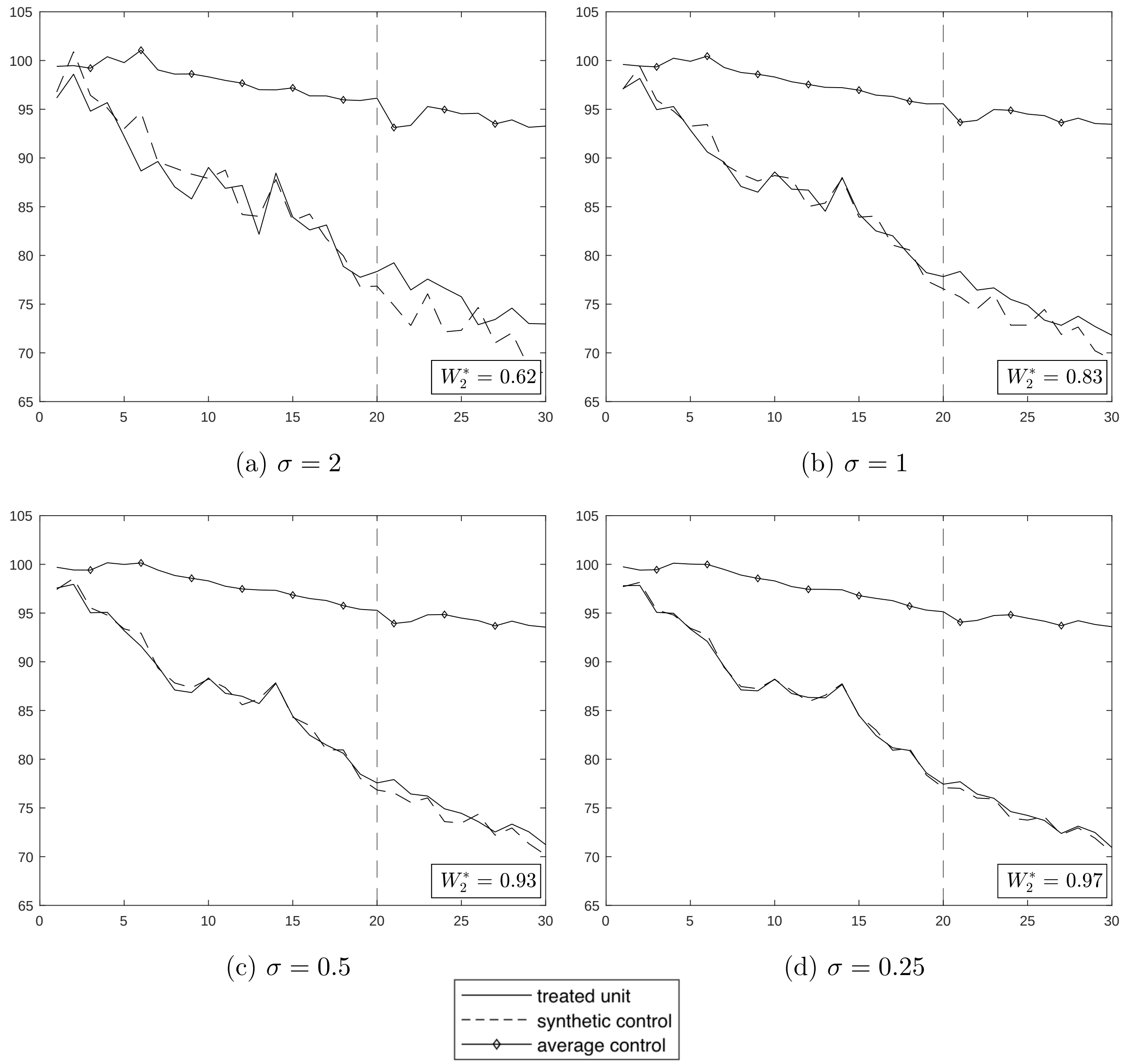
Note: 95% bands for the simulation design of Figure 4.2.

something to which the synthetic control method is able to adapt.

The result of Figure 4.4 is rather general. In a model like (4.2), non-stationary factor components help identify synthetic controls that reproduce the values of the μ_j for the treated. This is again the case in Figure 4.5, which repeats the analysis of Figure 4.2 but with $\rho = 1$, so the series λ_{ft} are non-stationary. As in Figure 4.4, the large variation in factor components cannot be compensated by variation in ϵ_{jt}. This results in minimal over-fitting in Figure 4.5, with W_2^* close to one for all values of σ in the figure.

Obtaining a good pre-treatment fit in the outcome variable (but also in other predictors of the outcome, as we will discuss later) is an important requisite for the performance of synthetic control estimators. However, good pre-treatment fit is not in itself a guarantee of good performance. The

Figure 4.4: Pre-treatment fit and estimation error with a stochastic trend



(a) $\sigma = 2$ (b) $\sigma = 1$

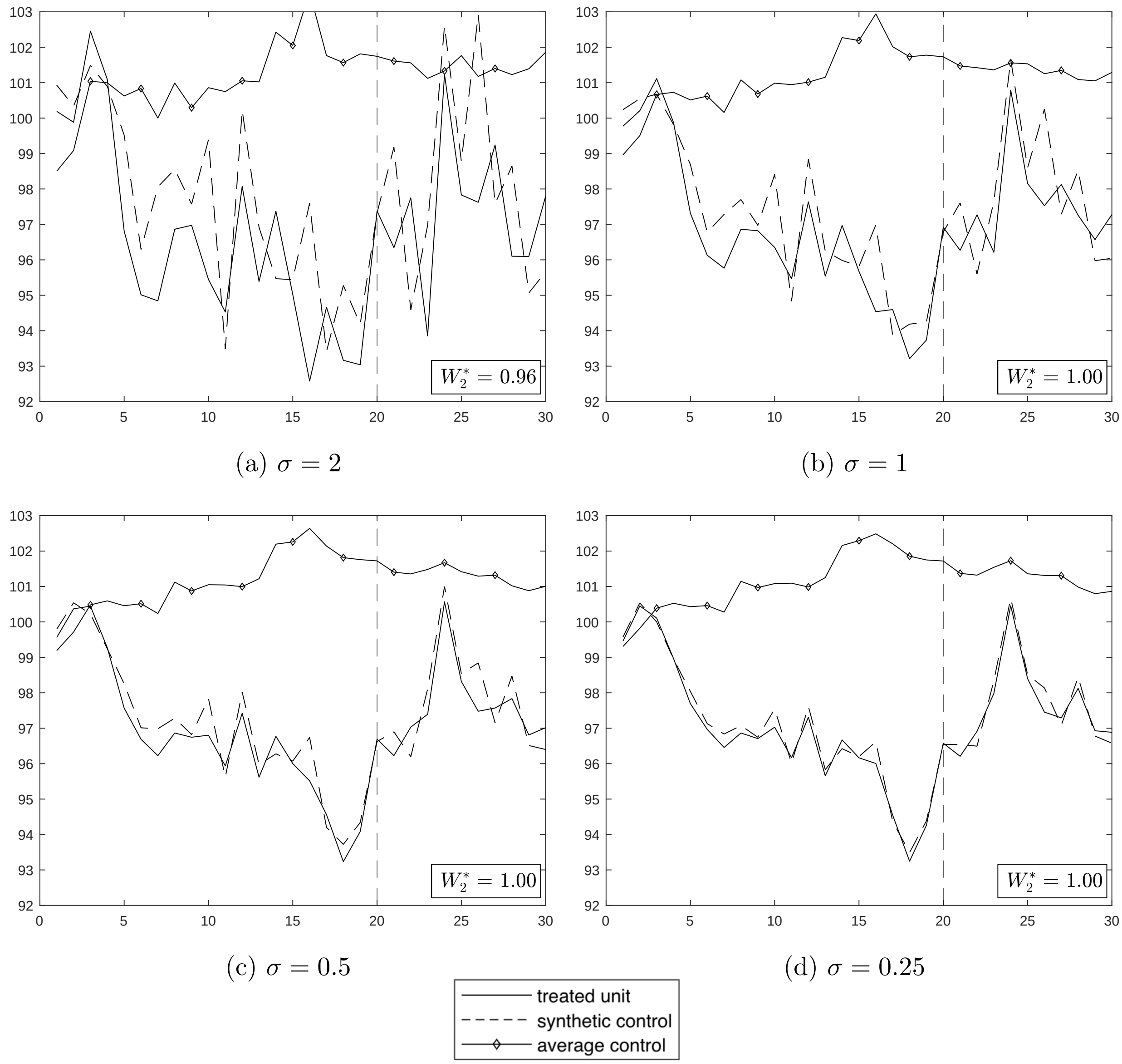
(c) $\sigma = 0.5$ (d) $\sigma = 0.25$

Note: Synthetic control estimates for the grouped factor model in equation (4.2), with $\rho = 0.5$, $T_0 = 20$, and $J + 1 = 20$ ($F = 10$). Relative to Figure 4.2, the series λ_{ft} incorporate stochastic trends. The average control assigns equal weights to all the untreated units.

reason is that, in some scenarios, a good pre-treatment fit may be obtained from variation in the individual transitory shocks, ϵ_{jt}, even when the selected synthetic control does not come close to reproducing the values of μ_j in equation (4.2) for the treated. This is what we have previously referred to as over-fitting.

A researcher employing synthetic controls should be able to identify settings conductive of over-fitting, and to modify the design of a study in order to avoid or attenuate over-fitting biases. We devote much of the rest of the article to discussing prevention, detection, and correction of

Figure 4.5: Pre-treatment fit and estimation error with $\rho = 1$



(a) $\sigma = 2$

(b) $\sigma = 1$

(c) $\sigma = 0.5$

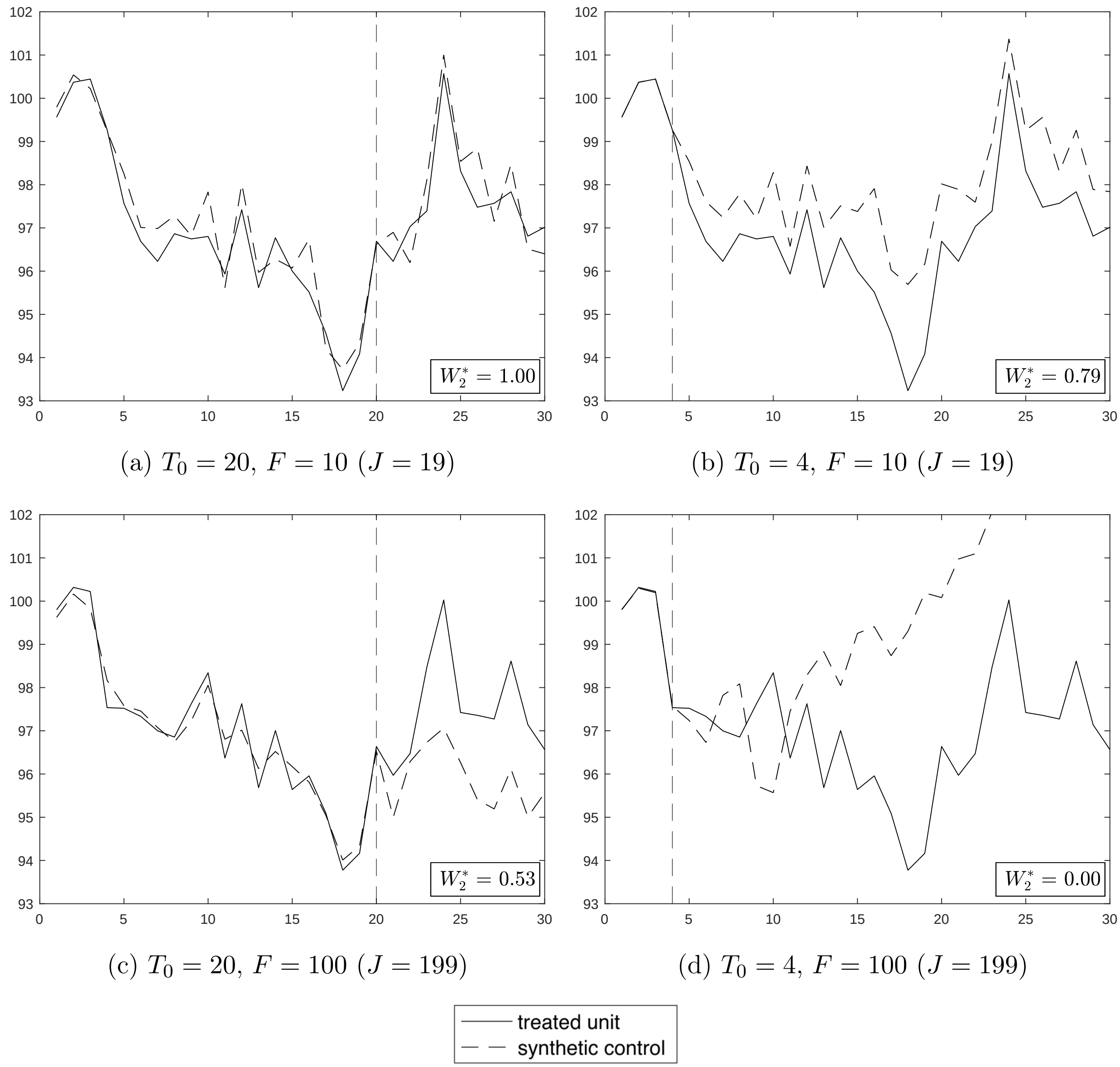
(d) $\sigma = 0.25$

Note: Synthetic control estimates for the grouped factor model in equation (4.2), with $\rho = 1$, $T_0 = 20$, and $J + 1 = 20$ ($F = 10$). Each panel reports results for a different value of σ. The average control assigns equal weights to all the untreated units.

over-fitted estimators. We discuss the scenarios where over-fitting is likely to bias synthetic control estimators and how to avoid those biases.

Figure 4.6 illustrates the effect of small T_0 and large J on over-fitting. Panel (a) of Figure 4.6 reproduces panel (c) in Figure 4.5, so $\rho = 1$ and $\sigma = 0.5$. In panel (a), the number of pre-treatment periods is $T_0 = 20$. Heterogeneity in the processes that govern the trajectories of the factors, λ_{ft}, identifies the ideal synthetic control with $W_2^* = 1$. That is, panel (a) depicts a setting with no over-fitting or bias. Variation in the transitory shocks, ϵ_{jt}, is equally well represented in the

Figure 4.6: Over-fitting with a short pre-treatment period and many donor units



(a) $T_0 = 20,\ F = 10\ (J = 19)$

(b) $T_0 = 4,\ F = 10\ (J = 19)$

(c) $T_0 = 20,\ F = 100\ (J = 199)$

(d) $T_0 = 4,\ F = 100\ (J = 199)$

Note: Synthetic control estimates for different values of T_0 and J, with $\rho = 1$ and $\sigma = 0.5$.

pre-treatment and post-treatment differences between the outcomes for the treated unit and the outcomes for the synthetic control. As we will see next, however, a small number of pre-intervention periods, T_0, or a larger number of donor pool units, J, may create over-fitting, breaking the close correspondence between pre-treatment fit and post-treatment error. Consider now panel (b), where the design of the simulation is the same as in panel (a) but with only $T_0 = 4$ pre-intervention periods. The fit in pre-intervention outcomes is now perfect. However, the small number of pre-intervention periods induces over-fitting, which creates large post-treatment estimation errors and a reduced role of the second sample unit in the synthetic control, with $W^* = 0.79$. Panel (c) goes back to $T_0 = 20$, as in panel (a), but increases the number of units in the donor pool to $J = 199$ (that is, $F = 100$). Similar to the effect of a small number of pre-treatment periods in panel (b), the increased number of units in the donor pool opens the door to over-fitting in panel (c). The post-treatment estimation error in panel (c) is much larger than in panel (a), even when the pre-treatment fit is better in panel (c). Moreover, the role of the second unit in the synthetic control is reduced to $W_2^* = 0.53$. Finally, panel (d) depicts the setting with $T_0 = 4$ and $J = 199$. As in panel (b), the pre-treatment fit is perfect. However, post-treatment estimation error is very large and the second unit does not contribute to the synthetic control, $W_2^* = 0$. On the whole, Figure 4.6 illustrates how the number of pre-treatment periods and the number of untreated units crucially affect the bias of the synthetic control estimator.

Apart from the possibility of over-fitting bias created by an excessively large J, preventing interpolation bias is another good reason to place restrictions on the untreated units that are allowed to enter the donor pool. Consider the factor model in equation (4.2). Even if ϵ_{jt} lacks variation and, therefore, over-fitting bias is not a concern, the linearity assumption in equation (4.2) may only be a local approximation. If that is the case, reproducing the predictor values for the treated unit by averaging untreated units that are far from the treated unit in the space of the predictors may result in interpolation biases. To attenuate interpolation biases, it is useful to restrict the units allowed in the donor pool to untreated units with similar values in the predictors, X_j, as the treated unit.

As we have discussed above, researchers applying synthetic control estimators should recognize settings conducive to biases—because of small values of T_0, large values of J, or because the units in the donor pool have values of the predictor that are substantially different from those of the treated unit—and adapt their designs to ameliorate bias concerns. In Section 4.4, we will discuss validation techniques to assess the potential for bias in concrete empirical settings.

We finish this section with a discussion of how modifications of the synthetic control design may affect the performance of the estimates. Panel (a) of Figure 4.7 plots the outcome values for a treated unit and its synthetic control, for a simulation design that is the same as for panel (d) or Figure 4.2 except that $T_0 = 15$. Because this is a simple setting without covariates, the synthetic control estimate of Y_{1t}^N in panel (a) is

$$\sum_{j=2}^{J+1} W_j^* Y_{jt},$$

with $W_2^*, \ldots, W_{J+1}^*$ chosen to minimize

$$\sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} W_j Y_{jt} \right)^2$$

with respect to $W_2, \ldots, W_{J+1}$, subject to the constrain that the weights, $W_2, \ldots, W_{J+1}$, are non-negative and sum up to one. The pre-treatment root-mean-square error (pre-RMSE),

$$\left(\frac{1}{T_0}\sum_{t=1}^{T_0}\Big(Y_{1t}-\sum_{j=2}^{J+1}W_j^*Y_{jt}\Big)^2\right)^{1/2}, \tag{4.6}$$

is equal to 0.23. The post-treatment root-mean-square error (post-RMSE)–which is analogous to (4.6), but calculated for the post-treatment periods—is somewhat higher, 0.33. Still, in this example, pre-intervention fit comes reasonably close to estimation accuracy in the post-intervention periods. As we will see next, adding extra flexibility to the synthetic control estimator may break down the correspondence between pre-RMSE and post-RMSE. Consider now panel (b) of Figure 4.7. It reports the results of a synthetic control estimator with an added constant shift (Doudchenko and Imbens, 2016). That is, panel (b) plots the result of estimating of Y_{1t}^N as

$$\widehat{Y}_{1t}^N = \widehat{\alpha} + \sum_{j=2}^{J+1}\widehat{W}_jY_{jt}, \tag{4.7}$$

where $\widehat{\alpha}, \widehat{W}_2, \ldots, \widehat{W}_{J+1}$ minimize

$$\sum_{t=1}^{T_0}\Big(Y_{1t}-\alpha-\sum_{j=2}^{J+1}W_jY_{jt}\Big)^2 \tag{4.8}$$

with respect to $\alpha, W_2, \ldots, W_{J+1}$, subject to the constrain that the weights, $W_2, \ldots, W_{J+1}$, are non-negative and sum up to one. Introducing the constant shift α, as in (4.8) is equivalent to computing a synthetic control with the outcome variables measured in deviations with respect to their pre-treatment means

$$Y_{jt} - \frac{1}{T_0}\sum_{k=1}^{T_0}Y_{jk}.$$

The availability of the constant α increases the degrees of freedom of the synthetic control estimator. This extra flexibility could be useful in certain instances, especially when the trajectory of the outcome for the treated unit is extreme relative to the trajectory of the outcomes in the donor pool, so a good fit is not attainable on the levels of Y_{jt}, but may be attainable on the demeaned outcomes. It is good to remember, however, that additional flexibility increases the potential for over-fitting. In the example of Figure 4.7, introducing a constant shift in the estimate may seem to be warranted at first sight. After all, the pre-treatment trajectory of the synthetic control seems to over-estimate the values of the treated outcomes. Indeed, panel (b) shows that the inclusion of the constant shift noticeably improves pre-treatment fit, with pre-RMSE dropping to 0.14. However, post-RMSE increases to 0.44. Panel (c) allows for additional flexibility relaxing the constraints on the weights (Hsiao et al., 2012; Doudchenko and Imbens, 2016). A simple regression estimator of this model takes the same form as $\widehat{Y}_{1t}^N$ in equation (4.7), but with parameters $\widehat{\alpha}, \widehat{W}_2, \ldots, \widehat{W}_{J+1}$

that minimize (4.8) for unrestricted values of $\alpha, W_2, \ldots, W_{J+1}$. Panel (c) in Figure 4.7 plots the result. Now, pre-RMSE is equal to zero, the unrestricted regression perfectly fits the trajectory of the outcome variable for the treated in the pre-treatment periods. However, the post-RMSE is 0.54, indicating a larger out-of-sample estimation error than in panels (a) and (b).

The increased risk of over-fitting that comes from increased flexibility in synthetic control estimation is well-recognized in the literature. At least since Doudchenko and Imbens (2016), researchers extending synthetic controls beyond the convex hull of the predictors' values for donor pool units have often combined the extra flexibility in the parameters with some regularization device—usually, but not exclusively, in the form of L_1 and/or L_2 penalizations on the synthetic control parameters (see, in particular, Amjad et al., 2018; Agarwal et al., 2021a; Arkhangelsky et al., 2021; Athey et al., 2021; Ben-Michael et al., 2021; Chernozhukov et al., 2021; Doudchenko and Imbens, 2016).

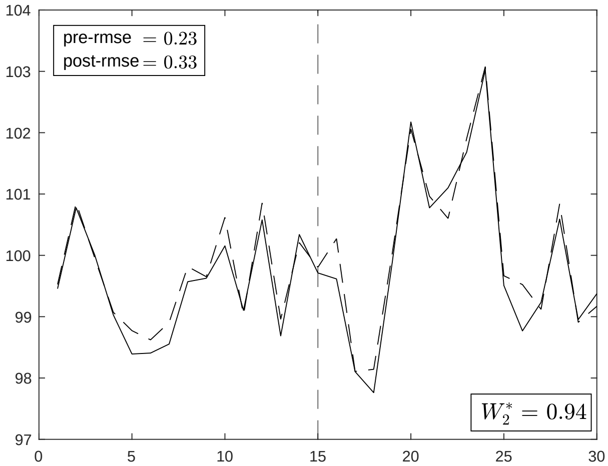
Figure 4.7 illustrates the effectiveness of restricting synthetic control weights to be non-negative and sum up to one as a regularization device. Relaxing the restrictions on the synthetic control weights, or adopting bias-correction techniques, as in Abadie and L'Hour (2021) and Ben-Michael et al. (2021), extends the applicability of the synthetic control method to settings where the values of the predictors for the treated unit cannot be closely fitted by a weighted average of the units in the donor pool. However, the increased flexibility in the synthetic control weights is obtained at the expense of the straightforward interpretability of the estimates. In addition, relaxing the restrictions on the synthetic control weights allows for unchecked extrapolation outside the support of the data, increasing the risk that the estimates reflect extreme counterfactuals, for which the data contain little information (King and Zeng, 2006).

4.4 Validating the Synthetic Control Estimator

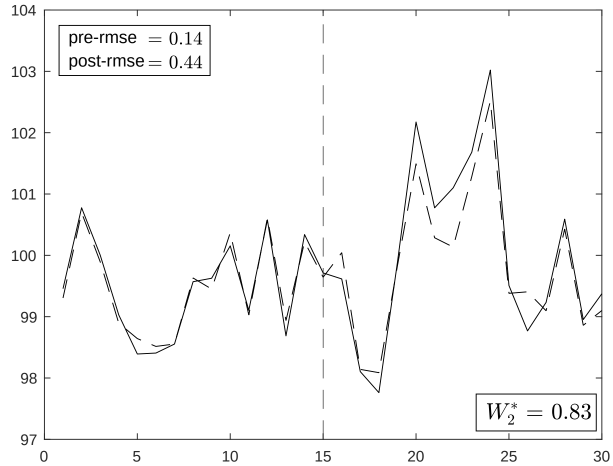
In Section 4.3 we have discussed how the scale of the transitory shocks, the number of pre-treatment periods, and the size of the donor pool influence the bias of the synthetic control estimator. In any particular empirical setting, however, a researcher could be uncertain about potential exposure to over-fitting and interpolation biases. In this section, we discuss validation techniques to help assess the magnitude of potential biases in concrete empirical applications.

An effective and visually interpretable validation exercise for synthetic controls can be obtained by artificially backdating the treatment of interest in the data. Figure 4.8 reports the results of backdating the treatment in the simulation setting of Figure 4.2. Recall that in this simulated scenario, the last pre-treatment period is $T_0 = 20$. In Figure 4.8, we have backdated the treatment so that the last pre-treatment period is now $T_0^b = 10$. This implies that we calculate the synthetic control weights using data for $t = 1, \ldots, T_0^b$ only. Backdating the treatment to $T_0^b = 10$ creates ten hold-out periods, $t = 11, \ldots, 20$, available to validate the predictions of the synthetic control estimator. Notice that this validation exercise could be performed at $t = T_0$, before post-treatment

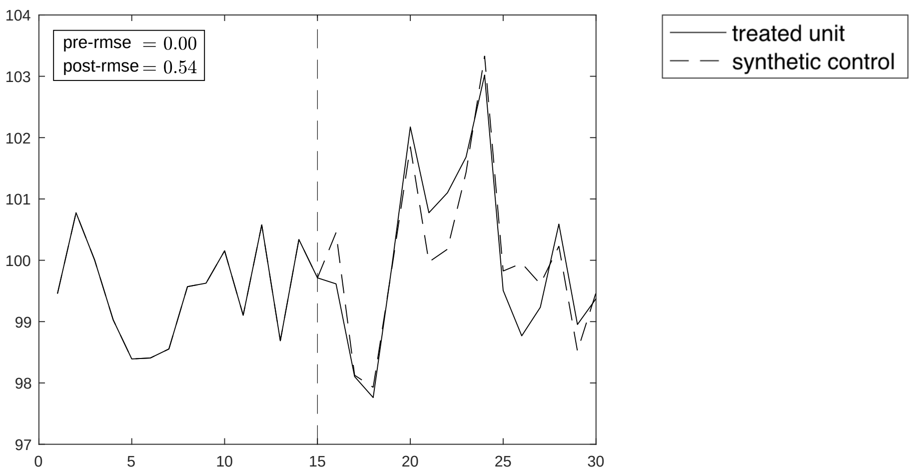
Figure 4.7: Over-fitting with added flexibility



(a) Synthetic control



(b) Synthetic control with constant shift



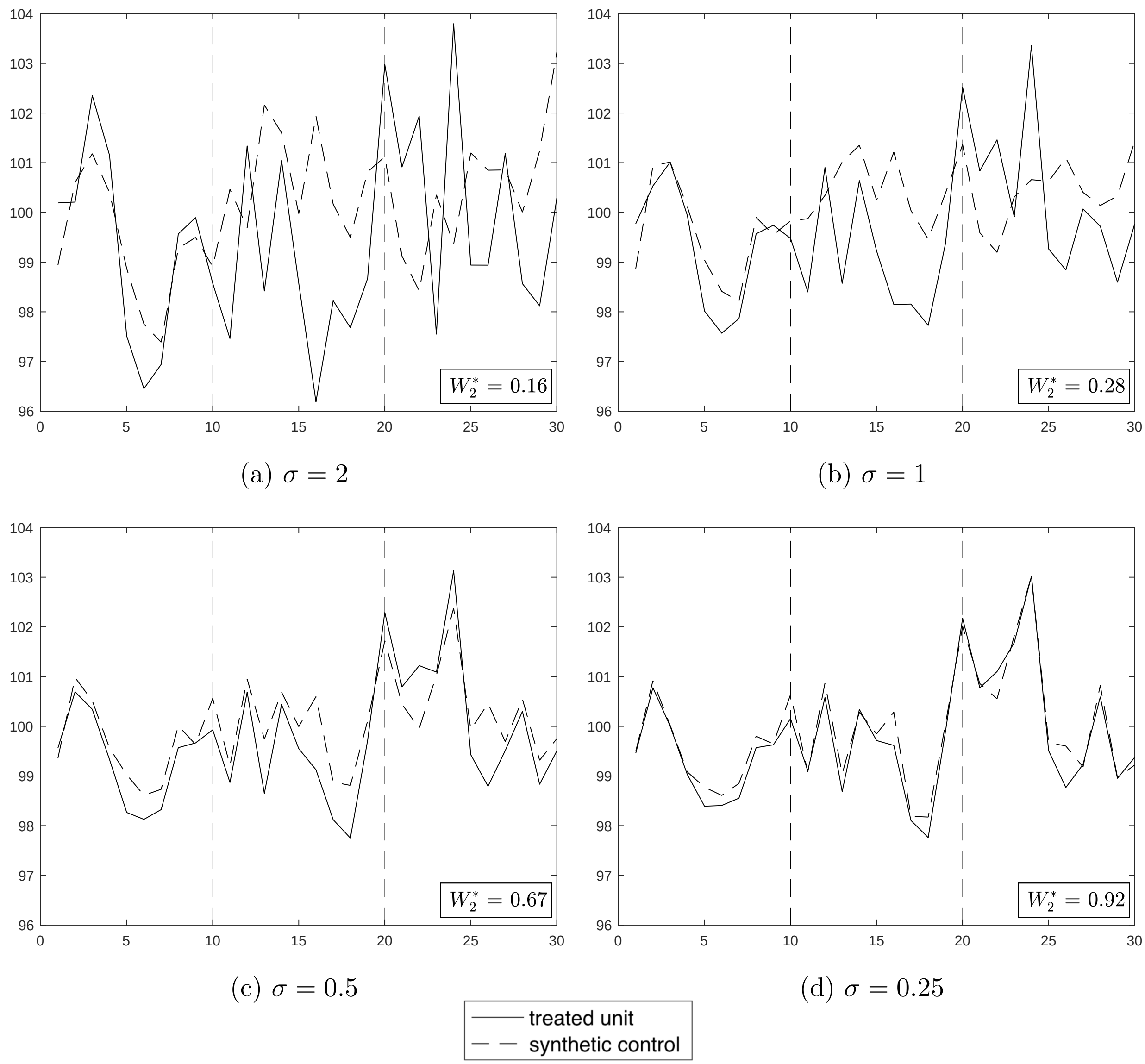
(c) Synthetic control via unrestricted regression

Note: Synthetic control estimates with added flexibility. Panel (a) reports results for the standard synthetic control estimator with weight constraints. Panel (b) reports results for a synthetic control model that includes a constant shift. Panel (c) reports results for a synthetic control estimated via unrestricted regression. $\sigma = 0.25$, $T_0 = 15$, $\rho = 0.5$ and $J = 19$.

data can possibly be observed, to inform study design decisions. An analyst carrying out this validation exercise at $t = T_0$ would observe large out-of-sample prediction errors for $\sigma = 2$ and $\sigma = 1$ in panels (a) and (b), but much more reliable predictions for $\sigma = 0.5$ and especially for $\sigma = 0.25$, in panels (c) and (d) respectively. In Figure 4.8, the quality of the prediction in the post-treatment periods, $t = 21, \ldots, 30$ closely mirrors the prediction error for the hold-out periods, $t = 11, \ldots, 20$.

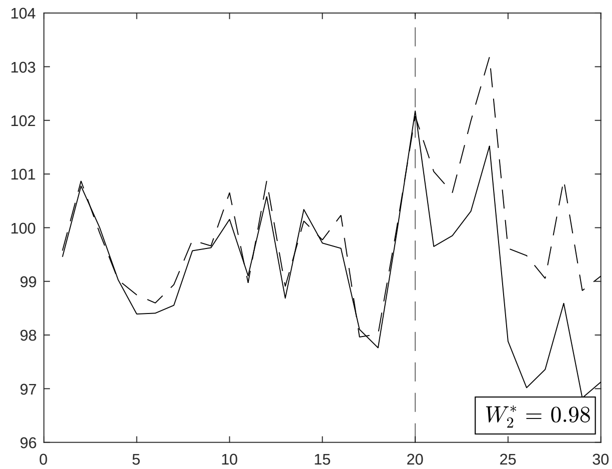
Figure 4.9 employs the same simulation design as in Figure 4.8, but with a positive treatment effect in the post-treatment periods. The effect of the treatment is detected in panel (a), where the synthetic control is trained using data up to period T_0. Panel (b) reports the result of the same backdating exercise as in Figure 4.8. Not only is the out-of-sample fit good during the hold-out periods, $t = 11, \dots, 20$, but the fact that an effect emerges at T_0, exactly the time when the true (in this case, simulated) treatment takes place and that the magnitudes of the estimates are similar in panels (a) and (b) provides additional credibility to the synthetic control estimates of panel (a).

Figure 4.8: Synthetic control validation



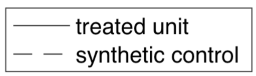
Note: Synthetic control estimates with treatment backdated to $T_0^b = 10$ for different values of σ. $T_0 = 20$, $\rho = 1$ and $J = 19$.

Figure 4.9: Backdating with a treatment effect.



(a) Intervention at $T_0 = 20$

(b) Backdating to $T_0^b = 10$



Note: Synthetic control estimates with a non-zero treatment effect and treatment backdated to $T_0^b = 10$. $\sigma = 0.25$, $T_0 = 20$, $\rho = 0.5$ and $J = 19$.

4.5 Trimming the Donor Pool

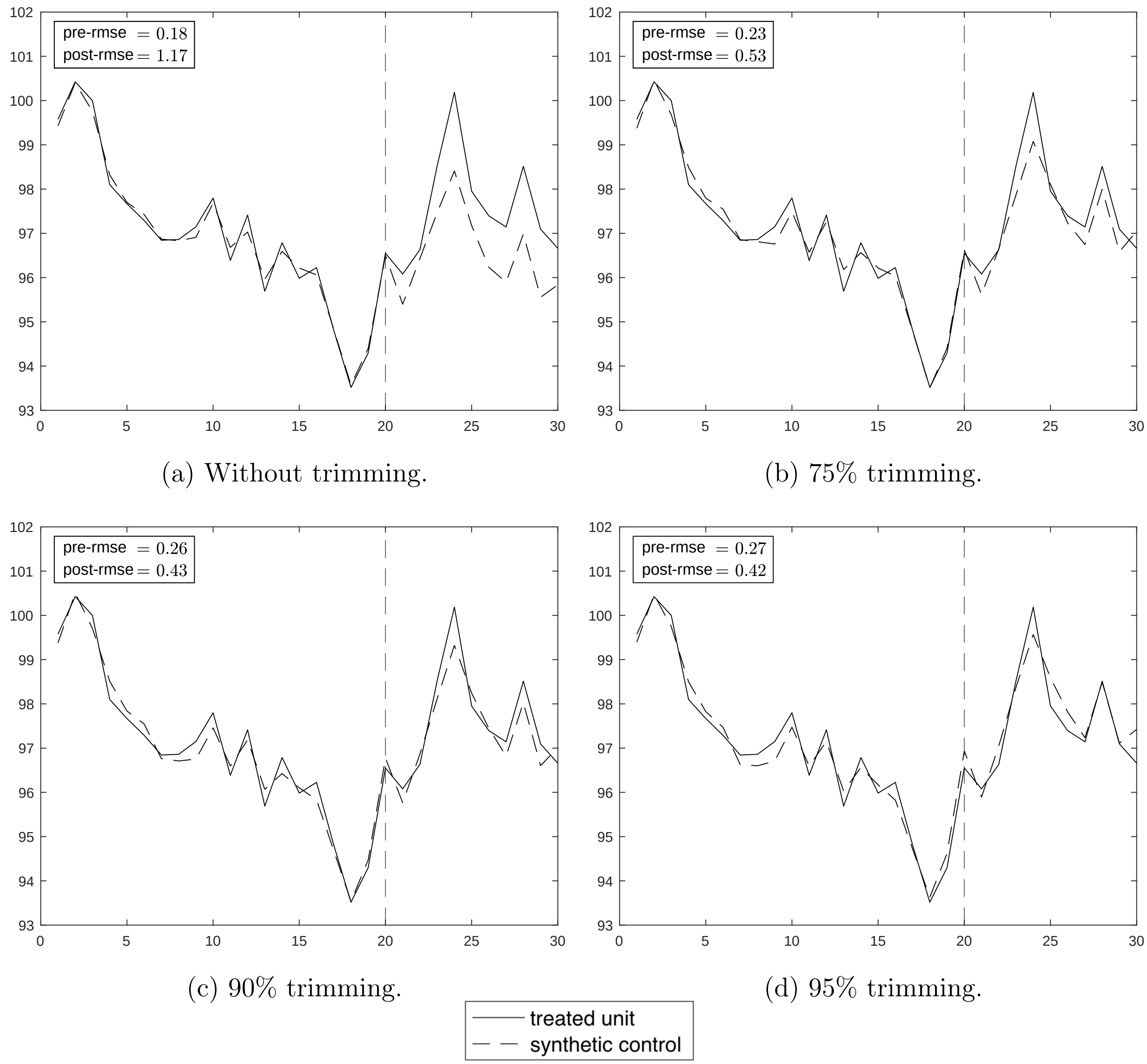
In Sections 4.2 and 4.3, we have discussed the importance controlling the size of the donor pool. In this section, we will show how actively trimming the donor pool to units with values of the predictors that are close to the values of the predictors for the treated unit can substantially improve the performance of synthetic control estimators.

The first panel of Figure 4.10 reports the result of a simulation based on the grouped factor model in equation (4.5), with $\rho = 1$, $\sigma = 0.25$, $T_0 = 20$, and $F = 100$ ($J = 199$). In this simulation, the large size of the donor pool promotes over-fitting resulting in large post-treatment estimation errors. In panels (b)-(d) of Figure 4.10 we have only retained the 50, 20, and 10 donor pool units, respectively, that best fit the values of the predictors for the treated (in this simple case, pre-treatment outcomes only). Trimming mechanically increases pre-treatment differences between the treated unit and the synthetic control, but noticeably improves the performance of the estimator in the post-treatment periods.

4.6 The Role of Observed Covariates

Thus far, we have only considered settings without covariates, $\boldsymbol{Z}_j$, in our simulations. In this section we discuss how the inclusion of observed covariates influences the performance of synthetic control

Figure 4.10: Synthetic control with trimming.



(a) Without trimming. (b) 75% trimming.

(c) 90% trimming. (d) 95% trimming.

Note: Synthetic control with and without trimming for $\sigma = 0.25$, $T_0 = 20$, $\rho = 1$ and $J = 199$. Panel (a) shows the standard synthetic control. Panel (b) shows the synthetic control with the donor pool trimmed to the 25% closest units to the treated unit according to the Euclidean distance. Panel (c) and (d) are like panel (b) with trimming to 10% and 5% of the units respectively.

estimators.

To illustrate the role of observed covariates, we consider the linear factor model in (4.2), with $\delta_t = 100$, $\boldsymbol{Z}_j$ and $\boldsymbol{\mu}_j$ uniformly distributed on $[0, 20]$, and where the components of $\boldsymbol{\theta}_t$ and $\boldsymbol{\lambda}_t$ follow independent random walks with standard Gaussian innovations. The idiosyncratic errors, ϵ_{jt}, are modelled as Gaussian noise with standard deviation equal to five. In Figure 4.11 we adopt a setting with $J = 1000$ and $T_0 = 4$ and five covariates. Panel (a) considers the case where the five covariates are unobserved (so $F = 5$) and there are no observed covariates, so a synthetic control estimator is based on pre-treatment outcomes only. The large value of J, the small value of T_0, and the presence of unobserved factors creates substantial over-fitting and estimation bias. In panels (b)-(f), we incrementally shift covariates from $\boldsymbol{\mu}_j$ to $\boldsymbol{Z}_j$, by including these covariates in the vectors X_j of predictive variables. That is, in contrast to panel (a), panels (b)-(f) use incrementally larger numbers of observed covariates, in addition to pre-treatment outcomes, to fit a synthetic control. In panel (f) there are no unobserved factors ($F = 0$): all five covariates are included in $\boldsymbol{Z}_j$ (and, therefore, in $\boldsymbol{X}_j$). The patterns of fit and estimation error in Figure 4.11 illustrate the importance of including predictive covariates in $\boldsymbol{X}_j$. Adding covariates to $\boldsymbol{X}_j$ induces a slight deterioration in pre-treatment fit in the outcomes, given that other covariates (apart from pre-treatment outcomes) enter $\boldsymbol{X}_j$. However, reductions in F are associated with substantial decreases in estimation error. As we increase the number of observed covariates, we rely less on pre-treatment outcomes as imperfect proxies for the values of the covariates, obtaining more reliable estimates. Figure 4.11 underscores the potential importance of including observed predictors of the outcomes, beyond pre-treatment outcomes, in the set of variables used to fit a synthetic control.

4.7 An Auto-Regressive Model

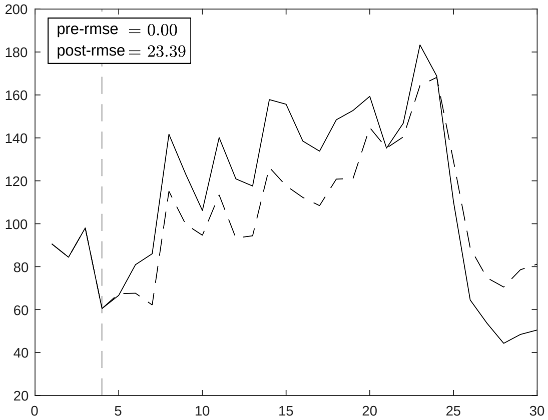
Although we have thus far used a linear factor structure in our simulations, it is important to notice that the validity of synthetic control estimators is not limited to the linear factor model in (4.2). In particular, Abadie et al. (2010) study the bias behavior of synthetic control estimators under a generative model with a vector auto-regressive structure in covariates and outcomes. In this section, we illustrate the applicability of synthetic control estimators under an auto-regressive data generating process. For that purpose, we consider a simple version, without covariates, of the auto-regressive model in Abadie et al. (2010),

$$Y_{jt}^N = \alpha_{jt} Y_{jt-1}^N + \epsilon_{jt},$$

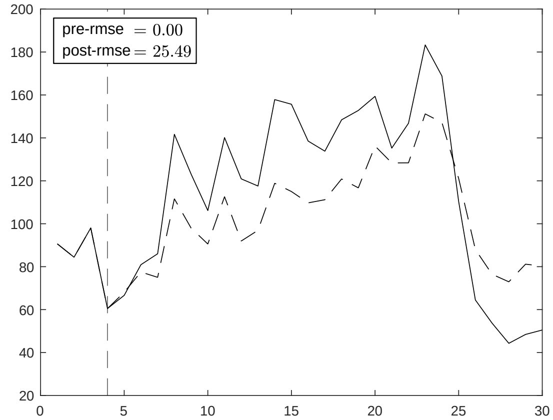
for $j = 1, \ldots, J+1$ and $t = 2, \ldots, T$, and $Y_{1t}^I = Y_{1t}^N$, for $t = T_0 + 1, \ldots, T$. This is an $AR(1)$ model with coefficients α_{jt} that are allowed to vary in time and across units. For the simulation results, we adopt $J + 1 = 50$ and, as before, $T_0 = 20$ and $T = 30$. We model Y_{j1}^N—the initial value of the outcome—for units $j = 1, \ldots J + 1$, as a Gaussian variable with mean 100 and standard deviation equal to 20. The error terms, ϵ_{jt}, are standard Gaussian and independent of each other, for $j = 1, \ldots J + 1$ and $t = 1, \ldots, T$. For α_{jt}, we adopt a grouped structure, with five groups of 10 units. Inside the groups, $\alpha_{j2}, \ldots, \alpha_{jT}$ are constant across units and drawn as independent Gaussian variables with mean equal to one and standard deviation equal to 0.1. The series $\alpha_{j2}, \ldots, \alpha_{jT}$ are independent across groups.

The result of a random simulation under this data generating process is reported in Figure 4.12. In this simulation, because of the assumed group-level heterogeneity in the trajectory of the outcome,

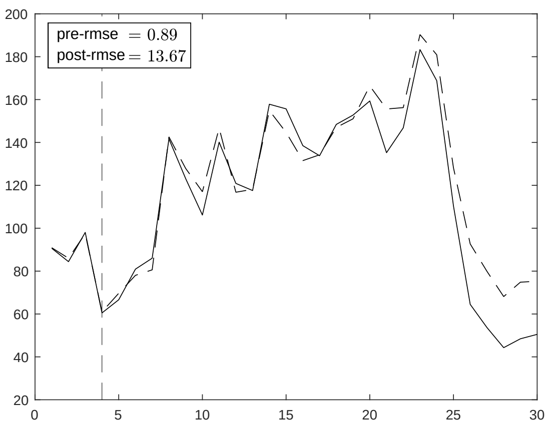
Figure 4.11: Including observed covariates



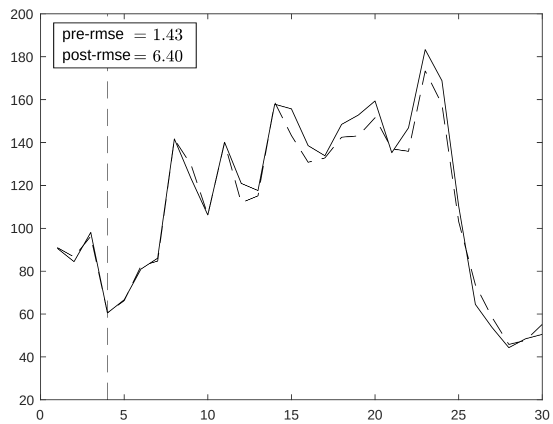
(a) $F = 5$



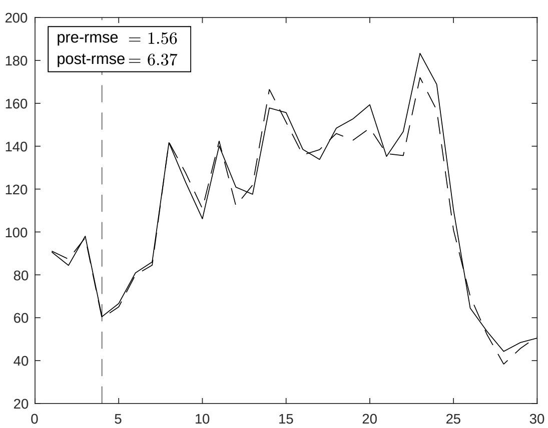
(b) $F = 4$



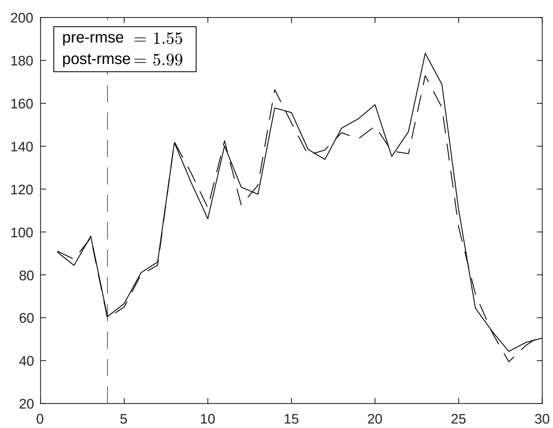
(c) $F = 3$



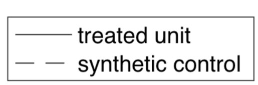
(d) $F = 2$



(e) $F = 1$



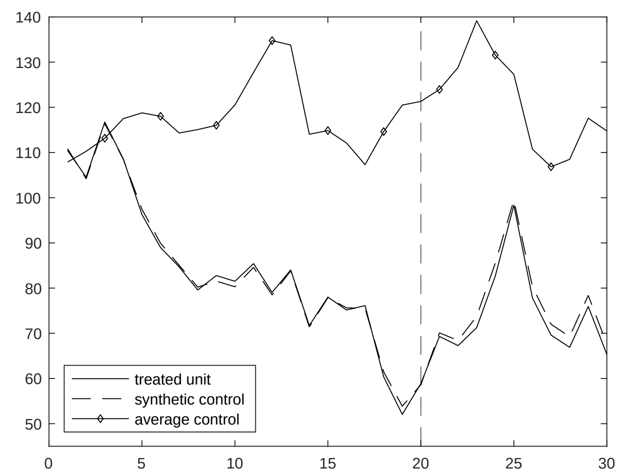
(f) $F = 0$



Note: Synthetic control under different values of F, the number of covariates not included in X_j. See main text for a detailed description of the data generating process.

the average of the outcome variable among all the untreated units is not able to reproduce the outcome for the treated unit, before or after the treatment is implemented. In contrast, the synthetic control estimates closely follow the trajectory of the outcome for the treated, demonstrating the applicability of synthetic control estimators outside the linear factor model considered in previous sections.

Figure 4.12: Synthetic control with an auto-regressive process.



Note: Synthetic control under an auto-regressive model for the outcome variable. $T_0 = 20$, $J + 1 = 50$. See text of Section 4.7 for a detailed description of the design of the simulation.

4.8 Conclusion

Synthetic controls are intuitive, transparent, and produce reliable estimates for a variety of data generating processes. However, like for any other statistical or econometric method to estimate treatment effects, there are settings where synthetic controls may fail. In this article, we have described the settings where the synthetic control method provides reliable estimates, and those where it does not. Of particular importance, we have discussed the nature of over-fitting in synthetic control estimators, when it arises, how to recognize it, and ways to alleviate or eliminate over-fitting biases.

Moreover, we have described intuitive and effective diagnosis checks for the validity of synthetic control estimates that can be easily carried out in empirical research. They pertain to the quality of the fit and length of the pre-intervention period, the ability to fit the treated unit using a relatively small donor pool of similar units, the ability to validate out-of-sample, and that a treatment effect estimate emerges at the time of the actual treatment implementation, even when the treatment is artificially backdated in the data.

Our recommendations are summarized in seven guiding principles for the research design of empirical studies that employ synthetic controls.

Appendix A

Appendix to Chapter 1

A.1 Theory

Throughout the appendix we introduce the following notation to refer to the debiased quantities: $\tilde{\epsilon}_{it} \equiv \epsilon_{it} - \sum_{j\neq i} \hat{w}_{ij}^{SC} \epsilon_{jt}$, as well as dropping the 'SC' weight subscript for notational convenience. Furthermore, we use T to mean T_1. The appendix consists of the following sections:

1. Bound on $\sum_{it} \tilde{Z}_{it}\tilde{\epsilon}_{it}$.
2. Proof of Theorem 1.
3. Proof of Theorem 2.
4. Proof of Theorem 3.
5. First stage debiasing lemmas.
6. Proof of Theorem 4.
7. Results for event study designs.
8. Results for projected and ensemble estimators.
9. Randomization inference.
10. Additional simulation tables.
11. Data appendix and additional tables for empirical examples.

A.1.1 Bound on $\sum_{it} \tilde{Z}_{it}\tilde{\epsilon}_{it}$

Lemma 2 (Bound on $\tilde{Z}_{it}\tilde{\epsilon}_{it}$). *Under Assumptions 1, 2, and 3, for any $\delta > 0$,*

$$\mathbb{P}\left(\left|\sum_{i,t>T_0} \tilde{Z}_{it}\tilde{\epsilon}_{it}\right| \geq \delta\right) \lesssim 2\exp\left(-\frac{\delta^2}{2c_z^2 JT\sigma_\epsilon^2}\right).$$

Hence, as $JT \to \infty$, $\frac{1}{JT}\sum_{it} \tilde{Z}_{it}\tilde{\epsilon}_{it} \xrightarrow{p} 0$.

Proof. First we show that the term has zero expectation given Assumption 3 and the independence of the error terms ϵ_{it}. The argument follows by noting that the SC weights depend only on ϵ_{it} for $t \leq T_0$,

$$w_j^{SC} \in \mathrm{argmin}_{w\in\mathcal{W}} \|Y_j^{T_0} - Y_{-j}^{T_0\prime} w\|^2,$$

as only data from the pre-treatment period is used, here denoted as the $(J+1)\times T_0$ matrix Y^{T_0}. Therefore, $w_j^{SC} \perp \epsilon_{it}$ for $t > T_0$. Recall that by the law of iterated expectations if random variables b is independent of z and a such that $\mathbb{E}[b|c] = 0$ a.e., then $\mathbb{E}[ab|z] = 0$. Using this fact, under

Assumption 2 it follows that $\mathbb{E}[\epsilon_{it} w_{ij}^{SC} | Z_{it}] = 0$ for $t > T_0$. Similarly, for any injective function $h : \text{Supp}(w) \to \mathbb{R}$ it follows that $h\left(w_j^{SC}\right) \perp \epsilon_{it}$ for $t > T_0$ and, consequently, $\mathbb{E}[\epsilon_{it} h\left(w_{ij}^{SC}\right) |Z_{it}] = 0$. To apply these facts, we re-write the second term, dropping the 'SC' subscript for convenience

$$\begin{aligned}
\mathbb{E}\left[\tilde{Z}_{it}\left(\epsilon_{it} - \sum_{j \neq i} \hat{w}_{ij}^{SC} \epsilon_{jt}\right)\right] &= \mathbb{E}\left[(Z_{it} - Z'_{-it} w_i)(\epsilon_{it} - \epsilon'_{-it} w_i)\right] \\
&= \mathbb{E}\left[Z_{it} \mathbb{E}[\epsilon_{it} - \epsilon'_{-it} w_i | Z_{it}] - Z'_{-it} \mathbb{E}[w_i(\epsilon_{it} - \epsilon'_{-it} w_i) | Z_{it}]\right],
\end{aligned}$$

where the $-i$ subscripts denote $J \times 1$ vectors not including unit i and w_i denote the $J \times 1$ vector of weights for unit i. Given that $\mathbb{E}[\epsilon_{it} w_{ij} | Z_{it}] = 0$ and $\mathbb{E}[\epsilon_{it} w_{ij}^2 | Z_{it}] = 0$, it follows that both conditional expectation terms are zero. Recall that for two vectors u and v, the following inequality holds $|u'v| \leq \|u\|_\infty \|v\|_1$. Therefore, Assumption 2, it follows that for $w \in \mathcal{W}$ for all i, t

$$|Z_{it} - Z'_{-it} w_i| \leq \|Z_t\|_\infty (1 + \|w_i\|_1) \leq c_z(1 + C),$$

where Z_t is the $J \times 1$ vector of instrument values at time t. It follows by the triangle inequality that

$$\left|\sum_{it} \tilde{Z}_{it} \tilde{\epsilon}_{it}\right| \leq (1 + C) c_z \left|\sum_{it} \tilde{\epsilon}_{it}\right|.$$

Given that ϵ_{it} are subgaussian and for $t > T_0$ independent of w_i, it follows that $\epsilon_{it} - \sum_{j \neq i} \hat{w}_{ij}^{SC} \epsilon_{jt}$ is a linear combination of subgaussian random variables. The first term has variance proxy σ_ϵ^2 and the second term has variance proxy $\sigma_\epsilon \|w_i\|^2 \leq C^2 \sigma^2$ as our weights satisfy $\|w_i\|^2 \leq \|w_i\|_1^2 \leq C^2$. Therefore, $\tilde{\epsilon}_{it}$ is subgaussian with variance proxy $2\sigma_\epsilon^2 C^2$. The result then follows directly by Hoeffding's inequality for subgaussian random variables (Theorem 2.6.2 Vershynin 2018)

$$\begin{aligned}
\mathbb{P}\left(\left|\sum_{it} \tilde{Z}_{it} \tilde{\epsilon}_{it}\right| \geq \delta\right) &\leq \mathbb{P}\left(\left|\sum_{it} \tilde{\epsilon}_{it}\right| \geq \delta / ((1 + C) c_z)\right) \\
&\lesssim 2 \exp\left(-\frac{\delta^2}{2C^2(1 + C)^2 c_z^2 J T \sigma_\epsilon^2}\right).
\end{aligned}$$

□

A.1.2 Proof of Theorem 1

Proof. We start by re-writing the factor structure in terms of the outcome variable and in the pre-treatment period

$$\tilde{\mu}_i' F_t = \tilde{Y}_{it} - \theta \tilde{R}_{it} - \tilde{\epsilon}_{it}.$$

Using the projection trick, we can rewrite $\tilde{\mu}_i$ in terms of pre-treatment quantities:

$$\tilde{\mu}_i = (F_{T_0} F'_{T_0})^{-1} F_{T_0} (\tilde{Y}_i^{T_0} - \theta \tilde{R}_i^{T_0} - \tilde{\epsilon}_i^{T_0}).$$

With this in mind, consider the object of interest

$$\begin{aligned}
\left| \sum_{it} \tilde{Z}_{it} \tilde{\mu}_i' F_t \right| &= \left| \sum_{it} \tilde{Z}_{it} F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} (\tilde{Y}_i^{T_0} - \theta \tilde{R}_i^{T_0} - \tilde{\epsilon}_i^{T_0}) \right| \\
&\leq \sum_{it} |\tilde{Z}_{it} F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} \tilde{\epsilon}_i^{T_0}| + \sum_{it} |\tilde{Z}_{it} F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} \tilde{Y}_i^{T_0}| + \theta \sum_{it} |\tilde{Z}_{it} F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} \tilde{R}_i^{T_0}| \\
&\leq (1+C) c_z \left(\sum_{it} |F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} \tilde{\epsilon}_i^{T_0}| + \sum_{it} |F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} \tilde{Y}_i^{T_0}| + \theta \sum_{it} |F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} \tilde{R}_i^{T_0}| \right).
\end{aligned}$$

Where the inequalities follow from the triangle inequality and the bound for $|\tilde{Z}_{it}| \leq (1+C) c_z$ which we derived in Lemma 2. For the first term bound we proceed as in Abadie and Vives-i-Bastida (2025) and apply the Cauchy-Schwartz inequality and the eigenvalue bound on the Rayleigh quotient to bound the factor terms for any t, s

$$(F_t' (F_{T_0} F'_{T_0})^{-1} F_s)^2 \leq \left(\frac{\bar{F}^2 k}{T_0 \underset{\sim}{\xi}} \right)^2,$$

To bound these terms in expectation observe that $\bar{\epsilon}_{it} \equiv F_t' (F_{T_0} F'_{T_0})^{-1} F_{T_0} \epsilon_{iT_0}$ is a linear combination of subgaussian random variables and therefore it is itself a subgaussian random variable with variance proxy $\left(\frac{\bar{F}^2 k}{T_0 \underset{\sim}{\xi}} \right)^2 \frac{\sigma_\epsilon^2}{T_0}$, where for notational convenience we use $\epsilon_{iT_0} \equiv \epsilon_i^{T_0}$. Therefore,

$$\begin{aligned}
|\mathbb{E}[(\bar{\epsilon}_{iT_0} - \bar{\epsilon}'_{-iT_0} w_i)]| &\leq \mathbb{E}[|(\bar{\epsilon}_{iT_0} - \bar{\epsilon}'_{-iT_0} w_i)|] \\
&\leq (1+C) \mathbb{E} \left[\sum_j |\bar{\epsilon}_{iT_0}| \right] \\
&\leq (1+C) \left(\mathbb{E} \left[\sum_j |\bar{\epsilon}_{iT_0}|^2 \right] \right)^2 \\
&= (1+C) \left(\sum_j \mathbb{E} \left[|\bar{\epsilon}_{iT_0}|^2 \right] \right)^2 \\
&\leq 2(1+C) \left(\frac{\bar{F}^2 k}{\underset{\sim}{\xi}} \right) \sqrt{\frac{J}{T_0}} \sigma_\epsilon.
\end{aligned}$$

The first inequality follows from Jensen's inequality. The second inequality follows by the triangle inequality, the absolute value and expectation operator inequality and $\|w_i\| \leq C$. The third follows from Holder's inequality with $q = 2$ and Jensen's inequality. Finally, the last inequality follows from Rigollet and Hutter 2019 (Lemma 1.4) which bounds absolute moments of sub-gaussian random variables. It follows that

$$\begin{aligned}\mathbb{E}[|F_t'(F_{T_0}F_{T_0}')^{-1}F_{T_0}(\epsilon_{iT_0} - \epsilon_{-iT_0}'w_i)|] &= \mathbb{E}[|(\bar{\epsilon}_{iT_0} - \bar{\epsilon}_{-iT_0}'w_i)|] \\ &\leq 2(1+C)\left(\frac{\bar{F}^2 k}{\underline{\xi}}\right)\sqrt{\frac{J}{T_0}}\sigma_\epsilon.\end{aligned}$$

For the term involving $\tilde{Y}$ we get that

$$\begin{aligned}\sum_{it} |F_t'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{Y}_i^{T_0}| &\leq \left(\frac{\bar{F}^2 k}{\underline{\eta} T_0}\right)\sum_{it}|\sum_{t\leq T_0}\tilde{Y}_{it}| \\ &= \frac{T}{T_0}\left(\frac{\bar{F}^2 k}{\underline{\eta}}\right)\sum_{i,t\leq T_0}|\tilde{Y}_{it}|\end{aligned}$$

Dividing by TJ it follows that the term is bounded by

$$\left(\frac{\bar{F}^2 k}{\underline{\eta}}\right)\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}_{it}| = \left(\frac{\bar{F}^2 k}{\underline{\eta}}\right)MAD(\tilde{Y}^{T_0}).$$

where $MAD(\tilde{Y}^{T_0}) = \frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}_{it}|$. An equivalent bound can be derived for the term involving $\tilde{R}$ which will depend on $\theta MAD(\tilde{R}^{T_0})$. The bound then follows from the proof of Theorem 1 and by Jensen's inequality applied to the absolute value. Consistency would follow by an application of Markov's inequality if the MAD terms are approximately zero. □

A.1.3 Proof of Theorem 2

Proof. The proof follows the proof of Theorem 1 by bounding $\mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}_{jt}|\right]$. It is useful to re-write the SC problem in matrix form. Let W be the $J \times J$ matrix of weights from program (1.3) where each row's sum is bounded by C and $diag(W) = 0$. Then the $J \times T_0$ matrix $\tilde{Y}^{T_0}$ can be re-written as $Y^{T_0} - \hat{W}Y^{T_0}$. It follows that the Frobenius norm of the matrix $\|\tilde{Y}^{T_0}\|_F^2 = \sum_{it<T_0}\tilde{Y}_{it}^2$ is bounded as follows

$$\begin{aligned}\|\tilde{Y}^{T_0}\|_F^2 = \|Y^{T_0} - \hat{W}Y^{T_0}\|_F^2 &\leq \|Y^{T_0}\|_F^2 + \|\hat{W}Y^{T_0}\|_F^2 \\ &\leq \|Y^{T_0}\|_F^2 + \|\hat{W}\|_F^2\|Y^{T_0}\|_F^2 \\ &\leq \|Y^{T_0}\|_F^2(1+C^2J) \\ &\leq \bar{r}_1\bar{\sigma}_1(1+C^2J).\end{aligned}$$

where the first inequality follows by the triangle inequality. The second by the bound on the frobenius norm of a matrix product. The third by noting that for $w_i \in \mathcal{W}$ each row of W summed is bounded by C and $\|w_i\|^2 \leq \|w_i\|_1 \leq C^2$ which implies $\|\hat{W}\|_F^2 \leq JC^2$. Finally, the last inequality follows from A4 as $\|Y^{T_0}\|_F^2 \leq \|Y^{T_0}\|_2^2\bar{r}$. Next, observe that $\sum_{it<T_0}|Y_{it}| = \|vec(Y^{T_0})\|_1^2$ and $\|vec(Y^{T_0})\|_2^2 = \|Y^{T_0}\|_F^2$.

So by the inequality between l_1 and l_2 norms,

$$\sum_{it \leq T_0} |Y_{it}| = \|vec(Y^{T_0})\|_1 \leq \sqrt{JT_0}\|vec(Y^{T_0})\|_2 = \sqrt{JT_0}\|Y^{T_0}\|_F.$$

Given the previous derivations we get the following bound

$$\frac{1}{JT_0}\sum_{i,t \leq T_0} |\tilde{Y}_{jt}| \leq \frac{\sqrt{JT_0}}{JT_0}\sqrt{\bar{r}_1\bar{\sigma}_1(1+C^2J)} \leq \sqrt{\bar{r}_1\bar{\sigma}_1}\left(\frac{1}{\sqrt{JT_0}} + C\sqrt{\frac{1}{T_0}}\right).$$

A similar derivation for the term involving $\tilde{R}$ yields

$$\frac{1}{JT_0}\sum_{i,t \leq T_0} |\tilde{R}_{jt}| \leq \frac{\sqrt{JT_0}}{JT_0}\sqrt{\bar{r}_2\bar{\sigma}_2(1+C^2J)} \leq \sqrt{\bar{r}_2\bar{\sigma}_2}\left(\frac{1}{\sqrt{JT_0}} + C\sqrt{\frac{1}{T_0}}\right).$$

The first part of the result then follows by noting that for a bounded random variable $|X| \leq K$, $\mathbb{E}|X| \leq K$. The consistency part follows by an application of Markov's inequality. □

A.1.4 Proof of Theorem 3

Lemma 3 (U projection consistency)**.** *Under the assumptions of Theorem 3 it follows that as $JT \to \infty$*

$$\frac{1}{JT}\|(I - \hat{W})U\|_1 \xrightarrow{p} 0,$$

where $\|\cdot\|_1$ denotes the sum of the absolute values elements of the matrix, I denotes the $J \times J$ identity matrix, $\hat{W}$ denotes the $J \times J$ matrix of weights from solving program 1.3 with $diag(\hat{W}) = 0$ and U denotes the $J \times T$ matrix of unobserved factors $\mu_i' F_t$. Furthermore, it follows that as $JT \to \infty$,

$$\frac{1}{JT}\|\hat{W} - P_U\|_F \xrightarrow{p} 0,$$

where $\|\cdot\|_F$ denotes the Frobenius norm, the sum of squared elements of the matrix, and P_U denotes the projection matrix $U(U'U)^{-1}U'$ for $J \times T$ matrix U.

Proof. From the proof of Theorem 1 we have that under the Assumptions 1-4 as $JT \to \infty$

$$\frac{1}{JT}\sum_{i,t>T_0} |\tilde{\mu}_i' F_t| \xrightarrow{p} 0.$$

Re-writing this statement in matrix form yields the first part of the proof. Let U be the $J \times T$ matrix with entries ${\mu_i}' F_t$ and $\hat{W}$ be the $J \times J$ matrix with zero diagonal elements and with each rows $\hat{w}_i^{*\prime}$, where $\hat{w}_i^*$ is the vector with entries equal to the weight vector $\hat{w}_i$ from program (1.3) with a zero in position i. Therefore, we have

$$\sum_{i,t>T_0} \tilde{\mu}_i' F_t = \sum_{i,t>T_0} [(I-\hat{W})U]_{it} = \|(I-\hat{W})U\|_1,$$

and given the results from Theorem 1 it follows that as $JT \to \infty$

$$\frac{1}{JT}\|(I-\hat{W})U\|_1 = \frac{1}{JT}\sum_{i,t>T_0} |\tilde{\mu}_i' F_t| \xrightarrow{p} 0,$$

which given that the LHS is non-negative and the RHS converges to zero in probability, implies the first part of the lemma.

For the second part, we are interested in the following object

$$\begin{aligned}\hat{W} - P_U &= \hat{W} - P_U + \hat{W}P_U - \hat{W}P_U \\ &= (I-P_U)\hat{W} - (I-\hat{W})P_U.\end{aligned}$$

By the triangle inequality we need to show that the following two objects on the LHS converge in probability to zero

$$\|\hat{W} - P_U\|_F \leq \|(I-\hat{W})P_U\|_F + \|(I-P_U)\hat{W}\|_F. \tag{A.1}$$

Define a linear operator $g_U : \mathbb{R}^{T\times J} \to \mathbb{R}^{J\times J}$ such that $g_U(A) := A(U'U)^{-1}U'$. Under Assumption 3 we have that $(U'U)^{-1}$ is invertible as the common factor matrix $F_T F_T'$ has bounded lowest eigenvalue and the factor loadings μ_i are bounded. Furthermore, given the bounds on the factor terms we have that g_U is a well defined bounded continuous linear operator. It follows by an application of the continuous mapping theorem for bounded functions that as $JT \to \infty$

$$\frac{1}{JT}\|(I-\hat{W})P_U\|_1 = \frac{1}{JT}\|g_U((I-\hat{W})U)\|_1 \xrightarrow{p} 0,$$

and convergence in the Frobenius norm follows from the l_1-l_2 inequality (for a vector a, $\|a\|_2 \leq \|a\|_1$). A similar argument can be applied to show that the second term in A.1 converges in probability, by noting that $P_U = P_U'$ and therefore that the Frobenius norm of both terms is the same. □

Proof. Under the assumptions, Lemma 2 and Theorem 3 show that both $\frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{\epsilon}_{it}$ and $\frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{\mu}_i' F_t$ are $o_p(1)$. It remains to be shown that the first stage term $\sum_{it}\tilde{R}_{it}\tilde{Z}_{it}$ does not go to zero in probability. Observe that

$$\begin{aligned}\frac{1}{JT}\sum_{it}\tilde{R}_{it}\tilde{Z}_{it} &= \frac{1}{JT}\sum_{it}(\gamma\tilde{Z}_{it} + \tilde{A}_{it} + \tilde{\eta}_{it})\tilde{Z}_{it} \\ &= \gamma\frac{1}{JT}\sum_{it}\tilde{Z}_{it}^2 + \frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{A}_{it} + \frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{\eta}_{it}.\end{aligned}$$

Under Assumptions 1-3 we have that $\frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{\eta}_{it} = o_p(1)$ by Lemma 2, given that the same assumptions for ϵ_{it} apply to η_{it}. Similarly, given $Z_{it} \perp A_{it}$, that Z_{it} is bounded, and $\frac{1}{JT}\sum_{i,t\geq T_0} A_{it} \xrightarrow{p} 0$ as $JT \to \infty$, it follows that $\frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{A}_{it} = o_p(1)$. Next, we show that the first term does not vanish in probability. Consider a $T \times J$ matrix Z for the instrument Z_{it} and an equivalent matrix $\tilde{Z}$ for the debiased instruments $\tilde{Z}_{it}$. We can express $\tilde{Z}$ in terms of Z and $\hat{W}$, the $J \times J$ matrix with zero diagonal elements and with each row $\hat{w}_i^{*\prime}$, where $\hat{w}_i^*$ is the vector with entries equal to the weight vector $\hat{w}_i$ from program (1.3) with a zero in position i,

$$\tilde{Z} = (I - \hat{W})Z = (I - \hat{W} + P_U - P_U)Z = (I - P_U)Z + (P_U - \hat{W})Z.$$

Recall that by the triangle inequality, for a norm $\|\cdot\|$ and matrices A, B it follows that $\|A + B\| \geq |\|A\| - \|B\||$. Using this inequality we have that

$$\begin{aligned}
\|\tilde{Z}\|_F &= \|(I - P_U)Z + (P_U - \hat{W})Z\|_F \\
&\geq |\|(I - P_U)Z\|_F - \|(P_U - \hat{W})Z\|_F| \\
&\geq \|(I - P_U)Z\|_F - \|(P_U - \hat{W})Z\|_F \\
&\geq \|(I - P_U)Z\|_F - \|(P_U - \hat{W})\|_\infty \|Z\|_1.
\end{aligned}$$

where the last inequality follows from the generalized Holder inequality for matrices, with the norms respectively representing maximum row and column sums. Given that $|Z_{it}| \leq c_z$ for all i,t it follows that $\|Z\|_1 \leq Jc_z$. By Lemma 3 we have that as $JT \to \infty$

$$\frac{1}{JT}\|(P_U - \hat{W})\|_1 \xrightarrow{p} 0, \tag{A.2}$$

and given that P_U is symmetric it follows that the same is true for the infinity norm $\|\cdot\|_\infty$, an note that given Lemma 3 it is sufficient to have that $T, T_0 \to \infty$ for fixed J. Therefore, we have that as $T, T_0 \to \infty$ $\frac{1}{JT}\|(P_U - \hat{W})\|_F^2 \|Z\|_F^2 \leq \frac{c_z}{T}\|(P_U - \hat{W})\|_1 \xrightarrow{p} 0$. For the term involving P_U, let $M_U = I - U_{JT_1}(U'_{JT_1}U_{JT_1})^{-1}U'_{JT_1}$ and $Z_{JT} = vec(Z)$, then

$$Z'_{JT}M_U Z_{JT} = Z'_{JT}M'_U M_U Z_{JT} = (M_U Z_{JT})'(M_U Z_{JT}) = \|M_U Z_{JT}\|_2^2 = \|(I - P_U)Z\|_F^2,$$

as M_U is idempotent. Combining (A.2) with $\frac{1}{JT}Z'_{JT}M_U Z_{JT} \xrightarrow{p} Q > 0$ as $JT \to \infty$ from Assumption 4, we have that as $JT \to \infty$

$$\frac{1}{JT}\sum_{it}\tilde{Z}_{it}^2 = \frac{1}{JT}\|\tilde{Z}\|_F^2 \geq Q - o_p(1).$$

It follows that this term is bounded below in probability. A similar argument yields that it is also bounded above in probability, since by the triangle inequality

$$\frac{1}{JT}\|\tilde{Z}\|_F^2 = \frac{1}{JT}\|(I-P_U)Z + (P_U - \hat{W})Z\|_F^2 \tag{A.3}$$

$$\leq \frac{1}{JT}\|\|(I-P_U)Z\|_F^2 + \frac{1}{JT}\|(P_U - \hat{W})Z\|_F^2. \tag{A.4}$$

which converges in probability to Q as $JT \to \infty$ by the same argument as above. It follows by the dominated convergence theorem that $\frac{1}{JT}\|\tilde{Z}\|_F^2 \xrightarrow{p} Q$. Therefore, we have that the first stage terms are $O_p(1)$. Finally, consider the synthetic IV estimator decomposition

$$\begin{aligned}
\tilde{\theta}^{TSLS} &= \theta + \left(\frac{1}{JT}\sum_{i,t>T_0}\tilde{Z}_{it}\tilde{R}_{it}\right)^{-1}\frac{1}{JT}\sum_{i,t>T_0}\tilde{Z}_{it}\left(\mu_i - \sum_{j\neq i}\hat{w}_{ij}^{SC}\mu_j\right)' F_t \\
&\quad + \left(\frac{1}{JT}\sum_{i,t>T_0}\tilde{Z}_{it}\tilde{R}_{it}\right)^{-1}\frac{1}{JT}\sum_{i,t>T_0}\tilde{Z}_{it}\left(\epsilon_{it} - \sum_{j\neq i}\hat{w}_{ij}^{SC}\epsilon_{jt}\right) \\
&= \theta + O_p(1)o_p(1) + O_p(1)o_p(1) \\
&= \theta + o_p(1)
\end{aligned}$$

Given that in the proof of Theorem 2 we bound away the contribution of the instrument $\tilde{Z}_{it}$, the same proof can be applied for Z_{it} for $\tilde{\theta}_{YR}^{TSLS}$. Furthermore, observe that

$$\tilde{Z}'Z = Z'(I-W)'Z = Z'(I-P_U)Z + Z'(P_U - \hat{W})'Z \tag{A.5}$$

and that because P_U is symmetric and idempotent, a corollary of Lemma 3 is that $\frac{1}{JT}\|\hat{W}' - P_U\|_1 \to 0$ in probability as $JT \to \infty$ since $\|\hat{W}' - P_U\|_F = \|(\hat{W} - P_U)'\|_F = \|\hat{W} - P_U\|_F$. Therefore, it follows that as $JT \to \infty$

$$\frac{1}{JT}\|(P_U - \hat{W})'Z\|_F^2 \xrightarrow{p} 0$$

by the same argument as for $\tilde{\theta}^{TSLS}$ and the same probability limit holds for the first stage term.

To show the consistency for $\tilde{\theta}_Z^{TSLS}$ observe that the equivalent decomposition (1.6) for the estimator involves bounding in probability the following term

$$\sum_{it}|\tilde{Z}_{it}\mu_i' F_t| = \sum_{it}|[Z'(I-\hat{W})'U]_{it}| = \|Z'(I-\hat{W})'U\|_1,$$

where $\|\cdot\|_1$ denotes the sum of the absolute value of the entries norm. Given that P_U is symmetric, a corollary of Lemma 3 is that $\frac{1}{JT}\|(I-\hat{W})'U\|_1 \to 0$ in probability as $JT \to \infty$ since $\|\hat{W}' - P_U\|_F = \|(\hat{W} - P_U)'\|_F = \|\hat{W} - P_U\|_F$. Therefore, by bounding Z_{it}, by Lemma 3 we have that $\frac{1}{JT}\sum_{it}|\tilde{Z}_{it}\mu_i' F_t| \xrightarrow{p} 0$ as $JT \to \infty$. The same argument as for $\tilde{\theta}_{YR}^{TSLS}$ for the first stage applies by noting that the same $\tilde{Z}_i t Z_{it}$ term appears. $\square$

A.1.5 Debiasing instrument Z

As noted in the main text debiasing the instrument is not a necessary condition for the consistent estimation of θ. In this section we note that while this is the case, debiasing the instrument can lead to better finite sample performance. In Lemma 4 we derive a probability bound for the unobserved factor term in decomposition (1.6) for the cases in which the instrument Z or $\tilde{Z}$ is used. The Lemma shows that both lead to the same rate, but with a different constant. If Z is used the rate is multiplied by $\frac{1}{JT}Z'Z \xrightarrow{p} Q_Z$, the variation in the instrument Z. While, if $\tilde{Z}$ is used the rate is multiplied by Q, as described in Assumption 3, the variation in the instrument once the unobserved confounder U is projected out. Given that $Q \leq Q_Z$ by defition we expect debiasing the instrument to lead to a better finite sample rate. On the other hand, one might think that debiasing the instrument leads to a worse first stage. While this may be true, it is not guaranteed that using $\tilde{Z}$ will make the first stage worse. Under additional assumptions Lemma 5 gives conditions under which using the instrument $\tilde{Z}$ leads to a weakly larger, but similar, first stage when we condition on the weights w. The intuition for the result is that $\tilde{Z}'Z$ appears in the first stage when the instrument Z is used versus $\tilde{Z}'\tilde{Z}$ in the case in which $\tilde{Z}$ is used. Suppose informally that $\tilde{Z} \simeq (I - P_U)Z$, where P_U is defined as in Assumption 3. Then, $\tilde{Z}'Z \simeq Z'(I - P_U)Z$ and $\tilde{Z}'\tilde{Z} \simeq Z'(I - P_U)'(I - P_U)Z = Z'(I - P_U)Z$ because $I - P_U$ is idempotent. Therefore, in both cases we do not expect the first stage to be very different.

In a simulation exercise, we plot the distribution of the empirical correlation between $\tilde{U}$, $\tilde{R}$ and Z and $\tilde{Z}$ in the context of our simulation design in section 1.6. As can be seen in Figure A.2 in the appendix simulation section, debiasing the instrument may matter a lot to reduce the finite sample correlation between the unobserved confounder and the instrument, but does not change the first stage significantly.

Lemma 4 (Probability bounds on $Z'\tilde{U}$, $\tilde{Z}'U$ and $\tilde{Z}'\tilde{U}$)**.** *Suppose in addition to Assumptions* $1-4$ *that* $\frac{1}{JT}Z'Z \xrightarrow{p} Q_Z$. *Then, under the assumptions of Theorem 4 it follows that as* $JT \to \infty$

$$\frac{1}{JT}\sum_{it}|Z_{it}\tilde{\mu}_i'F_t| \lesssim_P \sqrt{Q_Z} \times R(T, T_0, J),$$
$$\frac{1}{JT}\sum_{it}|\tilde{Z}_{it}\mu_i'F_t| \lesssim_P \sqrt{Q_Z} \times R(T, T_0, J),$$
$$\frac{1}{JT}\sum_{it}|\tilde{Z}_{it}\tilde{\mu}_i'F_t| \lesssim_P \sqrt{Q} \times R(T, T_0, J),$$

where $\lesssim_P$ *denotes bounded in probability and*

$$R(T, T_0, J) = \left(\frac{\bar{F}^2 kc}{\underline{\xi}}\right)\left(2c\frac{J}{\sqrt{T_0}}\sigma_\epsilon + (\bar{r}_1\bar{\sigma}_1 + \theta\bar{r}_2\bar{\sigma}_2)\left[\frac{1}{\sqrt{JT_0}} + JC^2\sqrt{\frac{T}{T_0}}\right]\right),$$

for $c = 1 + C$.

Proof. We start by noting that under the Assumptions of Theorem 2 we have that for the $JT \times 1$

vector $\tilde{U} = [\tilde{\mu}_i' F_t]_{it}$

$$\frac{1}{\sqrt{JT}} \mathbb{E}\|\tilde{U}\|_1 = \frac{1}{\sqrt{JT}} \sum_{i,t>T_0} \mathbb{E}|\tilde{\mu}_i' F_t| \leq R(T, T_0, J),$$

where the inequality follows from the proof of Theorem 4 as the proof uses a bounding argument to account for $\tilde{Z}$. Next, define the $JT \times 1$ vector of instruments $Z = [Z_{it}]_{it}$ and observe that by the Cauchy-Schwartz inequality and the l_1-l_2 norm inequality we have that

$$\frac{1}{JT}|Z'\tilde{U}| \leq \frac{1}{JT}\|Z\|_2\|\tilde{U}\|_2 \leq \frac{1}{\sqrt{JT}}\|Z\|_2 \frac{1}{\sqrt{JT}}\|\tilde{U}\|_1 \lesssim_P \frac{1}{\sqrt{JT}}\|Z\|_2 R(T, T_0, J),$$
$$\frac{1}{JT}|\tilde{Z}'\tilde{U}| \leq \frac{1}{\sqrt{JT}}\|\tilde{Z}\|_2 \frac{1}{\sqrt{JT}}\|\tilde{U}\|_2 \leq \frac{1}{\sqrt{JT}}\|\tilde{Z}\|_2 \frac{1}{\sqrt{JT}}\|\tilde{U}\|_1 \lesssim_P \frac{1}{\sqrt{JT}}\|\tilde{Z}\|_2 R(T, T_0, J),$$

where the last equality follows by applying Markov's inequality given that the expectation of the absolute value is bounded by rate R. The first result then follows by applying the continuous mapping theorem for the square root as $\frac{1}{JT}\|Z\|_2^2 \xrightarrow{p} Q_Z$. The second result follows by a similar argument to the proof of Theorem 2. As in the proof of Theorem 2, define the $T \times J$ matrix $\bar{Z}$, the $J \times J$ projection matrix U and the $J \times J$ weight matrix W. Then, we consider the following decomposition

$$\|\tilde{Z}\|_2 = \|(I - W)\bar{Z}\|_F = \|(I - P_U)Z + (P_U - \hat{W})Z\|_F^2.$$

In the proof of Theorem 2 we showed that $\frac{1}{JT}\|\tilde{Z}\|_F^2 \xrightarrow{p} Q$ as $JT \to \infty$. Therefore, by the continuous mapping theorem applied to the square root it follows that

$$\frac{1}{\sqrt{JT}}\|\tilde{Z}\|_2 \xrightarrow{p} \sqrt{Q},$$

which combined with the bound shows the first part of the proof. For the case of $\tilde{Z}'U$, we show that it is equivalent to the $Z'\tilde{U}$ case. We have that

$$\sum_{it} |\tilde{Z}_{it}\mu_i' F_t| = \sum_{it} |[Z'(I - \hat{W})'U]_{it}| = \|Z'(I - \hat{W})'U\|_1,$$

where $\|\cdot\|_1$ denotes the sum of the absolute value of the entries norm. Given that P_U is symmetric, a corollary of Lemma 3 is that $\frac{1}{JT}\|(I - \hat{W})'U\|_1 \to 0$ in probability as $JT \to \infty$ since $\|\hat{W}' - P_U\|_F = \|(\hat{W} - P_U)'\|_F = \|\hat{W} - P_U\|_F$. Therefore, by Lemma 3 we have that $\frac{1}{JT}\sum_{it} |\tilde{Z}_{it}\mu_i' F_t| \xrightarrow{p} 0$ as $JT \to \infty$ with the same rate $R(T, T_0, J)$. $\square$

Lemma 5 (First stage debiasing)**.** *Suppose that $Z_{it} = S_{it} + \mu_i' F_t$ and $A_{it} = 0$, where S_{it} is an i.i.d random variable such that $S_{it} \perp U_{it}$ and $\mathbb{E}[S_{it}] = 0$ and $\mathbb{E}[S_{it}^2] = \sigma_S^2$. Then, under the conditions of*

Theorem 2, for $\mathcal{W} = \Delta_J$, conditional on the weights w, it follows that as $JT \to \infty$

$$\frac{1}{JT}\sum_{it} \tilde{R}_{it}\tilde{Z}_{it} \xrightarrow{p} \gamma\tilde{\xi},$$
$$\frac{1}{JT}\sum_{it} \tilde{R}_{it}Z_{it} \xrightarrow{p} \gamma\sigma_S^2,$$

where $\sigma_S^2 < \tilde{\xi} \leq 2\sigma_S^2$.

Proof. Under the assumptions, the first stage in the case in which the instrument is debiased and in the case in which it is not is given by, respectively

$$\frac{1}{JT}\sum_{it} \tilde{R}_{it}\tilde{Z}_{it} = \gamma\frac{1}{JT}\sum_{it}\tilde{S}_{it}^2 + o_p(1),$$
$$\frac{1}{JT}\sum_{it} \tilde{R}_{it}Z_{it} = \gamma\frac{1}{JT}\sum_{it}\tilde{S}_{it}S_{it} + o_p(1),$$

where the $o_p(1)$ terms are the terms involving $\tilde{Z}_{it}\tilde{\eta}_{it}$ and $Z_{it}\tilde{\eta}_{it}$ which converge to zero in probability by Lemma 2, and the terms involving $\tilde{\mu}_i' F_t$ which converge to zero in probability by Theorem 1. Furthermore, by taking expectations with respect to S_{it} conditional on the weights w_i we have that

$$\mathbb{E}\left[\frac{1}{JT}\sum_{it}\tilde{S}_{it}S_{it}\right] = \frac{1}{JT}\sum_{it}\mathbb{E}[S_{it}^2] = \sigma_S^2,$$
$$\mathbb{E}\left[\frac{1}{JT}\sum_{it}\tilde{S}_{it}^2\right] = \frac{1}{JT}\sum_{it}\mathbb{E}[S_{it}^2](1+\|w_i\|_2^2) = \sigma_A^2\frac{1}{JT}\sum_{it}(1+\|w_i\|_2^2).$$

Given that the weights are in $\mathcal{W} = \Delta_J$, the simplex, it follows that $1/(J-1) \leq \|w_i\|_2^2 \leq 1$, so for all J and T we have that $\sigma_S^2 < \sigma_S^2 \frac{1}{JT}\sum_{it}(1+\|w_i\|_2^2) \leq 2\sigma_S^2$. The result follows under an appropriate LLN. $\square$

A.1.6 Proof of Theorem 4

In this section we provide a proof of Theorem 4 and an additional discussion in which we show that we can relax the conditioning on the weights by using a martingale representation of matching estimators as in Abadie and Imbens (2012).

Proof. We are interested in the following quantity for a given set of weights w and instrument Z:

$$\begin{aligned}\frac{\sqrt{JT}(\tilde{\theta}^{TSLS}-\theta)}{v^w_{JT}} &= \left(\frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{R}_{it}\right)^{-1}\frac{1}{v^w_{JT}\sqrt{JT}}\sum_{it}\tilde{Z}_{it}\left(\mu_i-\sum_{j\neq i}\hat{w}^{SC}_{ij}\mu_j\right)' F_t \\ &\quad+\left(\frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{R}_{it}\right)^{-1}\frac{1}{v^w_{JT}\sqrt{JT}}\sum_{it}\tilde{Z}_{it}\left(\epsilon_{it}-\sum_{j\neq i}\hat{w}^{SC}_{ij}\epsilon_{jt}\right),\end{aligned} \tag{A.6}$$

where the conditional variance is given by $v^w_{JT}=\frac{1}{\sqrt{JT}}\sum_{it}var(\tilde{Z}_{it}\tilde{\epsilon}_{it}\mid Z,w)$. First, we show that the bias term depending on $\mu_i'F_t$ in ($A.6$) is $o_p(1)$. The argument is the same as in the consistency theorem, but with a different rate. Note that from the proof of the consistency theorem (Theorem 3) that under Assumptions 1-4 we have that

$$\begin{aligned}\sum_{it}|F_t'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{Y}_i^{T_0}| &\leq \left(\frac{\bar{F}^2k}{\underline{\eta}T_0}\right)\sum_{it}|\sum_{t<T_0}\tilde{Y}_{it}| \\ &= \frac{T}{T_0}\left(\frac{\bar{F}^2k}{\underline{\eta}}\right)\sum_{i,t<T_0}|\tilde{Y}_{it}|\end{aligned}$$

So, dividing by $\sqrt{JT}$ and using the bound on the pre-treatment mean absolute deviation as in Theorem 2,

$$\begin{aligned}\frac{1}{\sqrt{JT}}\sum_{it}|F_t'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{Y}_i^{T_0}| &\leq \frac{\sqrt{T}}{T_0\sqrt{J}}\left(\frac{\bar{F}^2k}{\underline{\eta}}\right)\sum_{i,t<T_0}|\tilde{Y}_{it}| \\ &\lesssim \frac{\sqrt{T}}{T_0\sqrt{J}}\left(\frac{\bar{F}^2k}{\underline{\eta}}\right)\sqrt{JT_0\bar{r}\bar{\sigma}_1(1+J)} \\ &\lesssim \sqrt{\frac{T}{T_0}(1+J)\bar{r}\bar{\sigma}_1}.\end{aligned}$$

Therefore, the first term is $o_p(1)$ when $\sqrt{\frac{T}{T_0}(1+J)\bar{r}\bar{\sigma}_1}\to 0$. Similar argument follows for the R_{it} term.

Next, consider the sum in the second term in ($A.6$) and re-write it as follows

$$\begin{aligned}\sum_{it}\tilde{Z}_{it}\tilde{\epsilon}_{it} &= \sum_{it}\tilde{Z}_{it}(\epsilon_{it}-\sum_{j\neq i}w_{ij}\epsilon_{jt}) \\ &= \sum_{it}\epsilon_{it}(\tilde{Z}_{it}-\sum_{j\neq i}\tilde{Z}_{jt}w_{ji}) \\ &= \sum_{it}\epsilon_{it}\tilde{\alpha}_{it},\end{aligned}$$

where $\tilde{\alpha}_{it}=\tilde{Z}_{it}-\sum_{j\neq i}\tilde{Z}_{jt}w_{ji}$. Let $X_{it}\equiv\epsilon_{it}\tilde{\alpha}_{it}$. Given Assumption 4, it follows that conditional on

Z and w, X_{it} is an i.i.d random variable with mean zero and variance $\sigma^2\tilde{\alpha}_{it}^2$. Indeed, we have that $\mathbb{E}[X_{it}|Z,w] = \tilde{\alpha}_{it}\mathbb{E}[\epsilon_{it}|Z,w] = 0$, and $\mathbb{V}(X_{it}|Z,w) = \mathbb{E}[\epsilon_{it}\tilde{\alpha}_{it}|Z,w] = \tilde{\alpha}_{it}^2\mathbb{E}[\epsilon_{it}|Z,w] = \tilde{\alpha}_{it}^2\sigma^2$. Next, let $s_{JT}^2 = \sum_{it} var(X_{it}|Z,w) = \sigma_\epsilon^2\|\tilde{\alpha}\|_2^2$ and consider the Lindeberg CLT condition for $\delta > 0$ conditional on Z and weights w

$$
\begin{aligned}
\frac{1}{s_{JT}^2}\sum_{it}\mathbb{E}[X_{it}^2\mathbf{1}\{|X_{it}| > \delta s_{JT}\}] &= \frac{1}{\sigma_\epsilon^2\|\tilde{\alpha}\|_2^2}\sum_{it}\mathbb{E}[\epsilon_{it}^2\tilde{\alpha}_{it}^2\mathbf{1}\{|\epsilon_{it}| > \delta\frac{s_{JT}}{|\tilde{\alpha}_{it}|}\}] \\
&\leq \frac{1}{\sigma_\epsilon^2\|\tilde{\alpha}\|_2^2}\sum_{it}\mathbb{E}[\epsilon_{it}^2\tilde{\alpha}_{it}^2\mathbf{1}\{|\epsilon_{it}| > \delta\frac{s_{JT}}{max_{it}|\tilde{\alpha}_{it}|}\}] \\
&= \frac{1}{\sigma_\epsilon^2\|\tilde{\alpha}\|_2^2}\sum_{it}\tilde{\alpha}_{it}^2\mathbb{E}[\epsilon_{it}^2\mathbf{1}\{|\epsilon_{it}| > \delta\frac{s_{JT}}{max_{it}|\tilde{\alpha}_{it}|}\}] \\
&= \frac{1}{\sigma_\epsilon^2}\mathbb{E}[\epsilon_{it}^2\mathbf{1}\{|\epsilon_{it}| > \delta\frac{s_{JT}}{max_{it}|\tilde{\alpha}_{it}|}\}].
\end{aligned}
$$

We start by showing that $\frac{max_{it}|\tilde{\alpha}_{it}|}{s_{JT}} \to 0$ as $JT \to \infty$ with high probability. Note that

$$
max_{it}|\tilde{\alpha}_{it}| \leq c_z(1+C)(1+\|w^i\|_1),
$$

where w^i denotes the vector of weights assigned to unit i across the $J-1$ synthetic controls for the other units. A simple bound on $\|w^i\|_1$ is given by $\|w^i\|_1 \leq CJ$, so $max_{it}|\tilde{\alpha}_{it}| \leq c_z(1+C)(1+CJ)$. For the denominator, we show that under our Assumptions as $JT \to \infty$

$$
\frac{1}{JT}\|\tilde{\alpha}\|_2^2 \overset{p}{\not\to} 0.
$$

We do so by re-writting the object of interest in matrix form for $J \times T$ matrices Z, $\tilde{Z}$ and $J \times J$ matrix of weights W. By the triangle inequality and the generalized Holder inequality for matrices, we have that

$$
\begin{aligned}
\|\tilde{\alpha}\|_2^2 &= \|\tilde{Z} - W'\tilde{Z}\|_F^2 \\
&= \|(I-W)Z - W'(I-W)Z\|_F^2 \\
&\geq \|(I-W)Z\|_F^2 - \|W'(I-W)Z\|_F^2 \\
&\geq \|(I-W)Z\|_F^2 - \|W'(I-W)\|_\infty^2\|Z\|_1^2.
\end{aligned}
$$

By the proof of Theorem 2 we have that $\frac{1}{JT}\|(I-W)Z\|_F^2 \xrightarrow{p} Q > 0$ as $JT \to \infty$. For the second term in the last inequality, note that the instruments are bounded, so we have that $\|Z\|_1 \leq Jc_z$. Let P_U denote the $J \times J$ projection matrix for unobserved confounder $J \times T$ matrix U. It follows that, $P_U(I - P_U) = 0$. Then, consider

$$
\|W'(I-W)\|_\infty^2 = \|W(I-W')\|_1^2 = \|W(I-W') - P_U(I-P_U)\|_1^2.
$$

By Lemma 3 we have that $\frac{1}{\sqrt{JT}}\|W-P_U\|_1 \xrightarrow{p} 0$ as $JT\to\infty$. It follows that $\frac{1}{JT}\|W(I-W')\|_1^2 = \|W(I-W')-P_U(I-P_U)\|_1^2 \xrightarrow{p} 0$. Therefore, as $J/T\to 0$ we have that $\frac{1}{JT}\|\tilde{\alpha}\|_2^2 \geq Q$ with high probability. It follows by an application of the continuous mapping theorem that with high probability,

$$\frac{max_{it}|\tilde{\alpha}_{it}|}{s_{JT}} \leq \frac{(1/\sqrt{JT})c_z(1+C)(1+CJ)}{1/\sqrt{JT}\sigma_\epsilon\|\tilde{\alpha}\|_2} \lesssim_P \sqrt{\frac{J}{T}}\frac{c_zC^2}{\sigma_\epsilon\sqrt{Q}}.$$

Therefore, as $\sqrt{J/T}\to 0$ we have that $\frac{max_{it}|\tilde{\alpha}_{it}|}{s_{JT}} \to 0$. In the case in which we condition on a weight sequence such that $\|w^i\|_1 \leq c_w$ for all i, it follows that $max_{it}|\tilde{\alpha}_{it}| \leq c_z(1+C)(1+c_w)$, so the requirement that $\sqrt{J/T}\to 0$ is no longer necessary and the rate holds as long as $JT\to\infty$.

The Lindeberg condition holds by the Dominated Convergence Theorem given that ϵ_{it} is $i.i.d$ and has bounded fourth moments. Therefore, we have that as $JT\to\infty$ and either $J/T\to\infty$ or for all i, $\|w^i\|_1 \leq c_w < \infty$,

$$\frac{1}{\sqrt{JT}v^w_{JT}}\sum_{it}\tilde{Z}_{it}\tilde{\epsilon}_{it} = \frac{1}{s_{JT}}\sum_{it}\tilde{Z}_{it}\tilde{\epsilon}_{it} \xrightarrow{d} N(0,1).$$

Combining this result with the first stage terms in (A.6) we have that under the same rates of convergence

$$\frac{\sqrt{JT}(\tilde{\theta}^{TSLS}-\theta)}{v^w_{JT}} \xrightarrow{d} Q^{-1}N(0,1),$$

where we used Slutsky's theorem and the fact that as $JT\to\infty$ $\frac{1}{JT}\sum_{it}\tilde{Z}_{it}\tilde{R}_{it} \xrightarrow{p} Q$. Define v^*_{JT} as the probability limit of the conditional variance estimate $v^w_{JT} = \frac{1}{\sqrt{JT}}\sum_{it} var(\tilde{Z}_{it}\tilde{\epsilon}_{it} \mid Z, w)$ as $JT\to\infty$. It follows by Slutsky's theorem that

$$\frac{\sqrt{JT}(\tilde{\theta}^{TSLS}-\theta)}{v_{JT}} \xrightarrow{d} N(0,1),$$

where $v_{JT} = Q^{-2}plim\frac{1}{\sqrt{JT}}\sum_{it} var(\tilde{Z}_{it}\tilde{\epsilon}_{it} \mid Z, w)$.

$\square$

Discussion of weight conditioning Note that conditioning on the weights and instrument is equivalent to conditining on the pre-treatment period outcomes and treatments. Hence, it is possible consider a martingale representation as Abadie and Imbens (2012) do for matching estimators in which the information set is the pre-treatment period outcomes. Suppose for simplicity that we are in a setting where $R_{it}=0$ for $t<T_0$ and $\mathcal{W}=\Delta_{J-1}$. Then, define the partial sums for a given time t as

$$S_{tJk} = \sum_{l=1}^{k}\tilde{Z}_{lt}\tilde{\epsilon}_{lt},$$

under our assumptions it follows that

$$\mathbb{E}[S_{tJk+1} \mid S_{tJ1}, \dots, S_{tJk}] = \mathbb{E}[\tilde{Z}_{lt}\tilde{\epsilon}_{lt} \mid S_{tJ1}, \dots, S_{tJk}] + S_{tJk} = S_{tJk},$$

where the condition expectation is zero given that conditional on the weights w under our error independence and partial instrument validity assumptions the instrument and the error term are uncorrelated when $t > T_0$ as shown in Lemma 1. Furthermore, define the martingale difference as

$$X_{tJk} = S_{tJk} - S_{tJk-1} = \tilde{Z}_{kt}\tilde{\epsilon}_{kt},$$

and the information set is given by the generated σ-algebra $\mathcal{F}_{tJk} = \sigma(\{Y_1^{T_0}, \dots, Y_{k-1}^{T_0}, Z_{1t}, \dots, Z_{k-1t}\})$ as the weights depend only on the outcome values in the pre-treatment period. Therefore, conditing on $\mathcal{F}_{tJk}$ is equivalent to conditiong on the weight vectors.

We can now apply the martingale CLT (Theorem 3.2, p. 59 from Hall and Heyde (1980)):

Theorem 1 (Martingale CLT)**.** *Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ be a zero-mean, square-integrable martingale array with differences X_{ni} and let η^2 be an a.s. finite random variable. Suppose (1) a Lindeberg condition, for all $\varepsilon > 0$:*

$$\sum_i E\left(X_{ni}^2 \mathbf{1}\left\{|X_{ni}| > \varepsilon\right\} | \mathcal{F}_{n,i-1}\right) \xrightarrow{p} 0,$$

(2):

$$V_{nk_n}^2 = \sum_i E\left(X_{ni}^2 | \mathcal{F}_{n,i-1}\right) \xrightarrow{p} \eta^2,$$

and (3) the σ-fields are nested $\mathcal{F}_{n,i} \subset \mathcal{F}_{n+1,i}$ for $1 \leq i \leq k_n, n \geq 1$. Then:

$$S_{nk_n} = \sum_i X_{ni} \xrightarrow{d} Z,$$

where the random variable Z has characteristic function $E \exp\left(-\frac{1}{2}\eta^2 t^2\right)$.

Condition (3) is easy to check given our definition of $\mathcal{F}_{tJk}$. We start with condition (1) and consider $\frac{1}{\sqrt{JT}} X_{tJk}$. We will show point-wise convergence. Note that conditional on $\mathcal{F}_{tJk}$ by applying Holders inequality,

$$\mathbb{E}\left[\frac{X_{tJk}^2}{JT} \mathbf{1}\left\{|X_{tJk}| > \varepsilon\right\}\right] \leq \frac{1}{JT} \mathbb{E}\left[X_{tJk}^4\right]^{1/2} P\left(\left\{|X_{tJk}| > \sqrt{JT}\varepsilon\right\}\right)^{1/2}.$$

The second term can be further bounded by applying Chebyshev's inequality and under Assumptions

1-3

$$P\left(\left\{|X_{tJk}| > \sqrt{JT}\varepsilon\right\}\right) \leq \frac{var(\tilde{\epsilon}_{kt}|w)c_z^2}{JT\varepsilon^2} \leq \frac{\sigma^2 c_z^2}{T\varepsilon^2},$$

where the conditional variance is bounded above by the sum of the J variances. The first expectation term can be bounded by noting that the instruments are bounded and under the assumption of bounded fourth moments of the error term

$$\begin{aligned}\mathbb{E}\left[X_{tJk}^4\right] &\leq c_z^4 \mathbb{E}\left[\tilde{\epsilon}_{kt}^4\right] \\ &\leq c_z^4 \mathbb{E}\left[\sum_i \epsilon_{it}^4\right] \\ &= c_z^4 J \mathbb{E}\left[\epsilon_{it}^4\right].\end{aligned}$$

Combining the two bounds we get that for $X_{tJk}/\sqrt{JT}$,

$$\sum_{kt} E\left((X_{tJk}/\sqrt{JT})^2 \mathbf{1}\left\{|X_{tJk}/\sqrt{JT}| > \varepsilon\right\} |\mathcal{F}_{tJk-1}\right) \leq JT\frac{\sigma\sqrt{m_4}c_z^3\sqrt{J}}{JT\sqrt{T}\varepsilon} \lesssim \sqrt{\frac{J}{T}},$$

where $\mathbb{E}\left[\epsilon_{it}^4\right] \leq m_4$. Hence, as $\sqrt{\frac{J}{T}} \to 0$ Lindeberg's condition (1) is satisfied point-wise. Next, note that the variance term in condition (2) is bounded in probability and not $o_p(1)$ by the proof of Theorem 4. Hence, all conditions for the martingale Lindeberg CLT are satisfied and the normality results extends to the case in which we do not directly condition on the weights.

A.1.7 Results for event study designs

To derive theoretical results for the event study designs using the SIV debiased values, we impose additional assumptions on the instrument and the reduced form equation.

Assumption 1 (Event study design)**.** *The outcome of interest follows*

$$Y_{it} = \theta_t Z_i + \mu_i' F_t + \epsilon_{it},$$

where θ_t is the time varying parameter of interest satisfying $\theta_t = 0$ for $t \leq T_0$ and the instrument satisfies $Z_i \perp \epsilon_{it}, \eta_{it}$, $|Z_i| \leq c_z$, for all i, t, and, as $J \to \infty$,

$$\frac{1}{J} Z'(I - P_U)Z \xrightarrow{p} Q_e > 0,$$

where Z is the $J \times 1$ vector containing Z_i for all i and $P_U = U(U'U)^{-1}U'$ is the projection matrix *for the $J \times k$ matrix of factor loadings U.*

Note, that while Assumption 1 abstracts away from the instrument equation of the triangular design in Assumption 1, it can directly be reconciled with our design. Consider plugging in the instrument equation from Assumption 1 into the outcome equation and setting $A_{it} = 0$ such that

$$Y_{it} = \theta(\gamma Z_{it} + \eta_{it}) + \mu_i' F_t + \epsilon_{it}.$$

Furthermore, suppose we have a factor instrument, as is the case in the event study designs, and $Z_{it} = Z_i \times H_t$ with $H_t = 0$ for $t \leq T_0$, then we have that $\theta_t \equiv \gamma\theta H_t$ and $\theta_t = 0$ for $t \leq T_0$. Furthermore, the error term satisfies that $\epsilon_{it} + \theta\eta_{it} \perp Z_i$ as required by Assumption 1.
Under Assumption 1 we will show that the least-squares estimator for θ_t based on equation

$$\tilde{Y}_{it} = \sum_{k \neq T_0} \theta_k (\mathbb{1}\{t = k\} \times \tilde{Z}_i) + \epsilon_{it}, \tag{A.7}$$

recovers the dynamic effect θ_t under the same additional assumptions as Theorem 2. The SIV least squares event study estimator solves the following program for $\theta \in [\theta_1, \ldots, \theta_{T_0-1}, \theta_{T_0+1}, \ldots, \theta_T]$,

$$\min_\theta \sum_{it} (\tilde{Y}_{it} - \sum_{k \neq T_0} \theta_k (\mathbb{1}\{t = k\} \times \tilde{Z}_i))^2.$$

Taking FOC with respect to θ_l for some $l \neq T_0$ we have that

$$-2 \sum_{it} (\tilde{Y}_{it} - \sum_{k \neq T_0} \theta_k (\mathbb{1}\{t = k\} \times \tilde{Z}_i)) \mathbb{1}\{t = l\} \tilde{Z}_i = 0$$

$$\sum_i (\tilde{Y}_{il} - \theta_l \tilde{Z}_i) \tilde{Z}_i = 0,$$

as $\mathbb{1}\{t = k\}\mathbb{1}\{t = l\} = 1$ if and only if $k = l$. Hence, the SIV event study estimator is given by

$$\tilde{\theta}_l = \left(\sum_i \tilde{Z}_i^2 \right)^{-1} \sum_i \tilde{Z}_i \tilde{Y}_{il}.$$

Similarly to the SIV estimator for θ, we can decompose the event study estimator as follows

$$\tilde{\theta}_l = \theta_l + \left(\sum_i \tilde{Z}_i^2 \right)^{-1} \sum_i \tilde{Z}_i \tilde{\mu}_i' F_l + \left(\sum_i \tilde{Z}_i^2 \right)^{-1} \sum_i \tilde{Z}_i \tilde{\epsilon}_{il}. \tag{A.8}$$

In what follows, we prove a consistency result (Theorem 2) and asymptotic normality result (Theorem 3) for the event study estimator under the similar conditions as the results in the main text (Theorem 2 and Theorem 4). The main difference is that our results for the event study estimator require the number of units J and pre-treatment periods T_0 to grow to infinity, whether the results in the main text are valid for fixed J.

Theorem 2 (Event study estimates consistency). *Under Assumption 1 and the additional assumptions of Theorem 2, it follows that as $J \to \infty$ and $\bar{r}_1\bar{\sigma}_1\sqrt{\frac{J}{T_0}} \to 0$*

$$\tilde{\theta}_l \xrightarrow{p} \theta_l$$

for $l \in \{1, \ldots, T_0 - 1, T_0 + 1, \ldots, T\}$.

Proof. Start by considering the terms in A.8 that involve $\tilde{\epsilon}_{il}$. Under Assumption 3 and $Z_i \perp \epsilon_{it}$ it follows by Lemma 2 for $t = l$ that as $J \to \infty$

$$\frac{1}{J}\sum_i \tilde{Z}_i \epsilon_{il} \xrightarrow{p} 0.$$

Next, we consider the term involving $\tilde{\mu}_i' F_l$. Observe that under Assumption 1 we have that for $t \leq T_0$

$$\tilde{Y}_{it} = \tilde{\mu}_i' F_t + \tilde{\epsilon}_{it},$$

so we proceed as in the proof of Theorem 1 by projecting out the common factors and applying the triangle inequality.

$$\begin{aligned}
\left|\sum_i \tilde{Z}_i \tilde{\mu}_i' F_l\right| &= \left|\sum_i \tilde{Z}_i F_l'(F_{T_0}F_{T_0}')^{-1}F_{T_0}(\tilde{Y}_i^{T_0} - \tilde{\epsilon}_{it})\right| \\
&\leq \sum_i |\tilde{Z}_i F_l'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{\epsilon}_i| + \sum_i |\tilde{Z}_i F_l'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{Y}_i^{T_0}| \\
&\leq (1+C)c_z\left(\sum_i |F_l'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{\epsilon}_i| + \sum_i |F_l'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{Y}_i^{T_0}|\right).
\end{aligned}$$

Applying the bounds from the proof of Theorem 1 we have that

$$\frac{1}{J}\sum_i \mathbb{E}[|F_l'(F_{T_0}F_{T_0}')^{-1}F_{T_0}(\epsilon_{iT_0} - \epsilon_{-iT_0}' w_i)|] \leq 2(1+C)\left(\frac{\bar{F}^2 k}{\underline{\xi}}\right)\sqrt{\frac{J}{T_0}}\sigma_\epsilon,$$

and for the term involving $\tilde{Y}$ we get that

$$\frac{1}{J}\sum_i \mathbb{E}|F_t'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{Y}_i^{T_0}| \leq \left(\frac{\bar{F}^2 k}{\underline{\eta}}\right)\mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0}|\tilde{Y}_{it}|\right].$$

Given that the bounds depend on the same objects as in the proof of Theorem 2, under the stable rank Assumption 4, it follows that as $J \to \infty$ and $\bar{r}_1\bar{\sigma}_1\sqrt{\frac{J}{T_0}} \to 0$

$$\frac{1}{J}\sum_i \tilde{Z}_i \tilde{\mu}_i{}' F_l \xrightarrow{p} 0.$$

Applying Lemma 3 for the $t = l$ case, it follows that $\frac{1}{J}\sum_i \tilde{Z}_i^2 \xrightarrow{p} Q_e > 0$ under the same regime as $J \to \infty$, which completes the consistency result. $\square$

Theorem 3 (Asymptotic normality for event study estimates)**.** *Under Assumption 8, and the additional assumptions of Theorem 2, conditional on weights w and instruments Z_i, if $\sqrt{\frac{1}{T_0}}(1+J)\bar{r}_1\bar{\sigma}_1 \to 0$ and $\frac{1}{\sqrt{J}}\max_i \sum_{j\neq i}|w_{ji}| \to 0$, then as $J \to \infty$*

$$\frac{\sqrt{J}(\tilde{\theta}_l - \theta_l)}{v_J} \xrightarrow{d} (Q_e)^{-1} \times N(0,1).$$

where $v_J^2 = \frac{1}{J}\sum_i var(\tilde{Z}_i\tilde{\epsilon}_i \mid Z, w) = \frac{1}{J}\sum_i \sigma_\epsilon^2 \tilde{\alpha}_i^2$ and $\tilde{\alpha}_i = \tilde{Z}_i - \sum_{j\neq i}\tilde{Z}_j w_{ji}$.

Proof. We start by decomposing the estimator of interest

$$\begin{aligned}
\frac{\sqrt{J}(\tilde{\theta}_l - \theta_l)}{v_J^w} &= \left(\frac{1}{J}\sum_i \tilde{Z}_i^2\right)^{-1} \frac{1}{v_{Jl}^w\sqrt{J}}\sum_i \tilde{Z}_i^2 \left(\mu_i - \sum_{j\neq i}\hat{w}_{ij}^{SC}\mu_j\right)' F_l \\
&+ \left(\frac{1}{J}\sum_i \tilde{Z}_i^2\right)^{-1} \frac{1}{v_J^w\sqrt{J}}\sum_i \tilde{Z}_i \left(\epsilon_i - \sum_{j\neq i}\hat{w}_{ij}^{SC}\epsilon_{jl}\right),
\end{aligned} \tag{A.9}$$

where $(v_J^w)^2 = \frac{1}{J}\sum_i var(\tilde{Z}_i\tilde{\epsilon}_i \mid Z, w)$. The consistency of the factor term follows from a similar bounding argument as in the proof of Theorem 2. Given Assumption 3, we have that

$$\begin{aligned}
\frac{1}{\sqrt{J}}\sum_i |F_t'(F_{T_0}F_{T_0}')^{-1}F_{T_0}\tilde{Y}_i^{T_0}| &\le \frac{1}{T_0\sqrt{J}}\left(\frac{\bar{F}^2 k}{\underline{\eta}}\right)\sum_{i,t<T_0}|\tilde{Y}_{it}| \\
&\le \frac{1}{T_0\sqrt{J}}\left(\frac{\bar{F}^2 k}{\underline{\eta}}\right)\sqrt{JT_0}\bar{r}\bar{\sigma}_1(1+J) \\
&\lesssim \sqrt{\frac{1}{T_0}}(1+J)\bar{r}\bar{\sigma}_1.
\end{aligned}$$

Given that under Assumption 1 as $J \to \infty$ we have that $\frac{1}{J}\sum_i \tilde{Z}_i^2 \xrightarrow{p} Q_e > 0$, it follows that the first term in A.9 is $o_p(1)$ when $\sqrt{\frac{1}{T_0}}(1+J)\bar{r}_1\bar{\sigma}_1 \to 0$ as $T_0, J \to \infty$.

We proceed by showing that the second term A.9 converges in distribution to a normal random variable.

$$\begin{aligned}
\sum_i \tilde{Z}_i\tilde{\epsilon}_{il} &= \sum_i \tilde{Z}_i(\epsilon_{il} - \sum_{j\neq i} w_{ij}\epsilon_{jl}) \\
&= \sum_i \epsilon_{il}(\tilde{Z}_i - \sum_{j\neq i}\tilde{Z}_j w_{ji}) \\
&= \sum_i \epsilon_{il}\tilde{\alpha}_i,
\end{aligned}$$

where $\tilde{\alpha}_i = \tilde{Z}_i - \sum_{j\neq i}\tilde{Z}_j w_{ji}$. As in the proof of Theorem 4, we show that the Lindeberg CLT condition holds as $J \to \infty$. Let $X_{il} \equiv \epsilon_{il}\tilde{\alpha}_i$. Given Assumption 1, it follows that conditional on Z and w, X_{il} is an i.i.d random variable with mean zero and variance $\sigma^2\tilde{\alpha}_i^2$. Indeed, we have that $\mathbb{E}[X_{il}|Z,w] = \tilde{\alpha}_i\mathbb{E}[\epsilon_{il}|Z,w] = 0$, and $\mathbb{V}(X_{il}|Z,w) = \mathbb{E}[\epsilon_{il}\tilde{\alpha}_i|Z,w] = \tilde{\alpha}_i^2\mathbb{E}[\epsilon_{il}|Z,w] = \tilde{\alpha}_i^2\sigma^2$. Next, let $s_J^2 = \sum_{it} var(X_{il}|Z,w) = \sigma_\epsilon^2\|\tilde{\alpha}\|_2^2$ and consider the Lindeberg CLT condition for $\delta > 0$ conditional on Z and weights w.

$$
\begin{aligned}
\frac{1}{s_J^2}\sum_i \mathbb{E}[X_{il}^2\mathbf{1}\{|X_{il}| > \delta s_J\}] &= \frac{1}{\sigma_\epsilon^2\|\tilde{\alpha}\|_2^2}\sum_i \mathbb{E}[\epsilon_{il}^2\tilde{\alpha}_i^2\mathbf{1}\{|\epsilon_{il}| > \delta\frac{s_J}{|\tilde{\alpha}_i|}\}] \\
&= \frac{1}{\sigma_\epsilon^2}\mathbb{E}[\epsilon_{il}^2\mathbf{1}\{|\epsilon_{il}| > \delta\frac{s_J}{\max_i |\tilde{\alpha}_i|}\}].
\end{aligned}
$$

Under our assumptions we have that $\frac{1}{\sqrt{J}}\max_i |\tilde{\alpha}_i| \to 0$ as $J \to \infty$. Therefore, it suffices to show that $\frac{1}{\sqrt{J}}s_J$ converges to a constant to guarantee the Lindeberg condition, given that we assume bounded fourth moments. A similar argument to the proof of Theorem 4 shows that with high probability as $J \to \infty$

$$
\frac{1}{J}\|\tilde{\alpha}\| \geq Q_e > 0,
$$

so by an application of of the continuous mapping theorem we have as $J \to \infty$,

$$
\frac{s_J}{\max_i |\tilde{\alpha}_i|} \to 0.
$$

The asymptotic normality result then follows by an application of Slutsky's Theorem and the Lindeberg CLT. □

Event study variance estimation The conditional variance v_J^2 can be estimated directly from the data using the plug-in estimator for idiosyncratic shock variance $\hat{\sigma}_\epsilon^2$. In our applications, we let the variance of the error term vary with l, and we estimate the variance as follows

$$
\tilde{\sigma}_{l,SIV}^2 = \frac{\hat{\sigma}_{\epsilon,l}^2\|\tilde{\alpha}\|_2^2}{\sum_i \tilde{Z}_i^2},
$$

where $\hat{\sigma}_{\epsilon,l}^2 = \frac{1}{J-1}\sum_i \hat{\epsilon}_{il}^2$, where $\hat{\epsilon}_{il}^2$ is the residual for the event study regression for time period parameter θ_l. Standard errors and asymptotically valid confidence intervals can be constructed using this variance estimator under the same assumptions as Theorem 3, if additionally $\mathbb{E}[\epsilon_{il}^4] < \infty$ for all l. In the event study figures, e.g. Figure 1.4, the 95% confidence intervals vary with time because we use the time-varying variance estimator.

A.1.8 Projected and ensemble estimators

The bound for the projected estimator can be derived in the same way as for the SIV estimator in the proof of Theorem 1. We give a sketch of the proof. The only difference for the projected estimator is that we now have that

$$\tilde{\mu}_i^P = (F_{T_0} F_{T_0}')^{-1} F_{T_0} (\tilde{Y}_i^{P,T_0} - \theta \tilde{R}_i^{P,T_0} - \tilde{\epsilon}_{it}^P).$$

Where the $\tilde{\epsilon}_{it}^P$ denotes the projection into the instrument space Z_i. Given that $Z_i \perp \epsilon_{it}$ and ϵ_{it} is mean zero, it will mean that with high probability the weights do not depend on the ϵ_{it} terms in the pre-treatment period. Hence, removing the contribution of the J other units in the rate.

Under Assumptions 1-4, the same bounds apply for $\mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0} |\tilde{Y}_{jt}^P|\right] + \theta \mathbb{E}\left[\frac{1}{JT_0}\sum_{i,t\leq T_0} |\tilde{R}_{jt}^P|\right]$ for the projected estimator than for the SIV up to an additional constant depending on Q (from the projection operator $Z'(Z'Z)^{-1}Z$). This highlights that in cases in which the instrument does not explain a lot of the variation in U then projecting into the instrument space makes matching U harder.

Therefore, the projected SIV is also a consistent estimator of θ under the same conditions as the SIV. Hence, for any $\alpha \in (0, 1)$ the ensemble estimator is consistent.

Aggregation estimator As an alternative to the projected estimator when Z_{it} does not follow a factor structure, we consider an aggregation estimator that matches on the aggregated timeseries in the pre-period. The Aggregation estimator is computed equivalently to the projected estimator except that the synthetic control weigths are computed as follows

1. Let $Q_i = \sum_{t<T_0} Z_i Y_{it}$ for each i.
2. Match the aggregated values for each i

$$w_j^A \in \text{argmin}_{w \in \Delta^{J-1}} \|Q_i - Q_{-i}' w\|^2$$

3. Compute the aggregated TSLS estimator $\tilde{\theta}^A$ using the w^A weights.

In the appendix additional simulation section we provide results for the aggregation estimator and show that it performs badly in settings in which the time series component is important.

A.1.9 Randomization Inference

An alternative to the permutation based test described in section 1.5 is a test based on randomization inference in the spirit of Imbens and Rosenbaum (2005). Instead of considering the differences in effects across time periods we now fix outcomes Y and consider different assignment distributions for the instrument-treatment pairs (R, Z) across units. The intuition is that under a uniform assignment distribution of (R, Z) we should not see an effect on the outcome Y. We can construct a permutation t-statistic and p-value using the following procedure.

1. In the pre-period compute the SC weights and generate the debiased quantities.

2. Define the set of permutations of the J units: $\mathcal{P}(J)$.

3. For a given permutation $\pi \in \mathcal{P}(J)$, compute

$$\tilde{\theta}_{\pi} = \left(\sum_{it} \tilde{Z}_{\pi(i)t} \tilde{R}_{\pi(i)t} \right)^{-1} \sum_{it} \tilde{Z}_{\pi(i)t} \tilde{Y}_{it},$$

where we permute the individuals for Z and R but not Y.

4. p-value:

$$\hat{p} = \frac{1}{\mathcal{P}(J)} \sum_{\pi \in \mathcal{P}(J)} P(\tilde{\theta}_{\pi} \geq \tilde{\theta}_{TSLS})),$$

where absolute values should be used for a two-sided test.

This test will be a valid test for the null $H_0 : \theta = 0$ under the assumption of exchangeability across units of $\{\mu_i, \epsilon_{it}, \eta_{it}$. The reason why we do not implement this test for our main empirical example is that we do not believe that μ_i, which is correlated with the distance share Z_i is exchangeable across units. It should also be noted that in general $\mathcal{P}(J)$ might be very large and in practice the p-values should be computed by randomly sampling from $\mathcal{P}(J)$.

A.1.10 Additional simulations

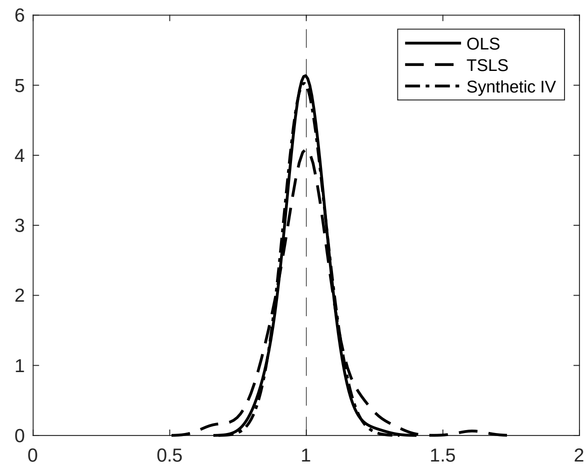
In this section, we consider the same simulation design as in section 1.6 but under different parameters. In particular, we consider a setting with a weaker instrument and a smaller signal variance. Figure A.1 replicates the first five panels of Figure 1.5 for a simulation design with parameters $\beta = \gamma = 1$, $k = 1$, $T = 30$, $T_0 = 20$, $J = 20$, $\sigma_\epsilon = 0.5$, $\kappa = 0.5$, $\sigma_\eta = \sigma_z = \sigma_g = 1$.

Table A.1 replicates Table 1.1 for the simulation design of Figure A.1 and Table A.3 for the design with $T_0 = 10$. In all cases we find that the SIV estimator, projected SIV and ensemble estimator outperform the OLS and TSLS estimators. Furthermore, the performance of the SIV whe the number of pre-treatment periods is halved remains good. Table A.2 evaluates the coverage of the 95% confidence intervals for the synthetic IV using $\hat{v}_{JT_1}$ of the true parameter $\theta = 1$ for different correlation settings and post-treatment periods. We find that in settings in which the OLS and TSLS are unbiased the synthetic IV exhibits a slight over-coverage, in the well behaved settings with moderate noise and correlation the coverage is good, and in high correlation settings, as expected, we report under-coverage.

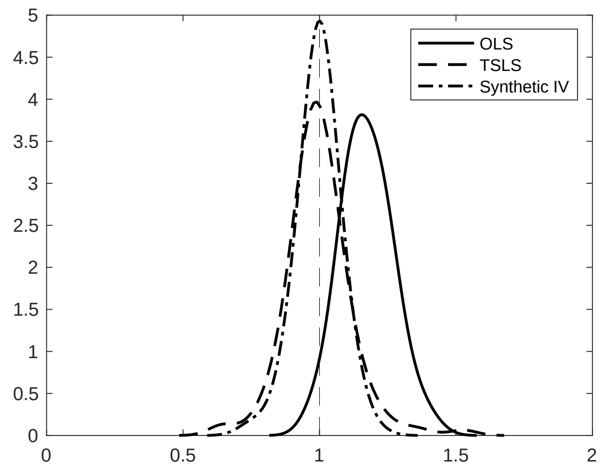
Simulations with comparison to factor model estimation

A natural question given our linear factor model design is how does the SIV compare to methods that estimate the factor structure directly. This question is one that applies not only to our setting, but to any SC design in which a linear factor model is used to derive statistical properties for the

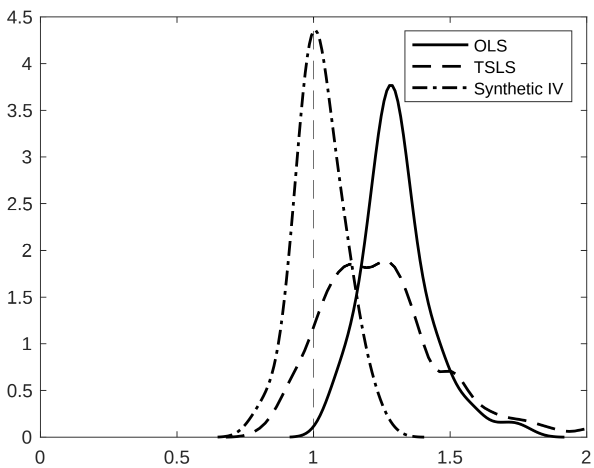
Figure A.1: Model comparison in simulations



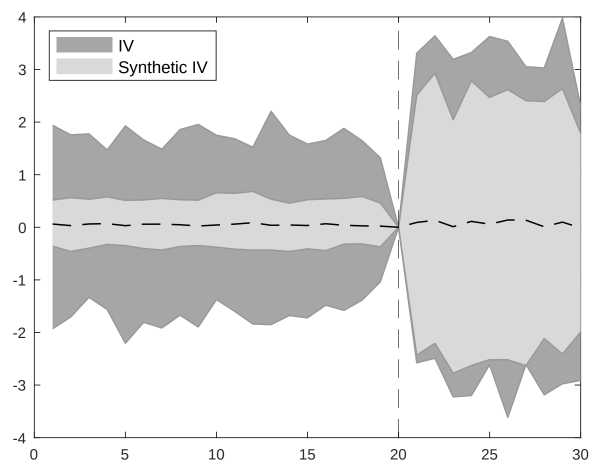
(a) $\rho = \rho_z = \rho_g = 0$



(b) $\rho = 0.5$, $\rho_z = \rho_g = 0$



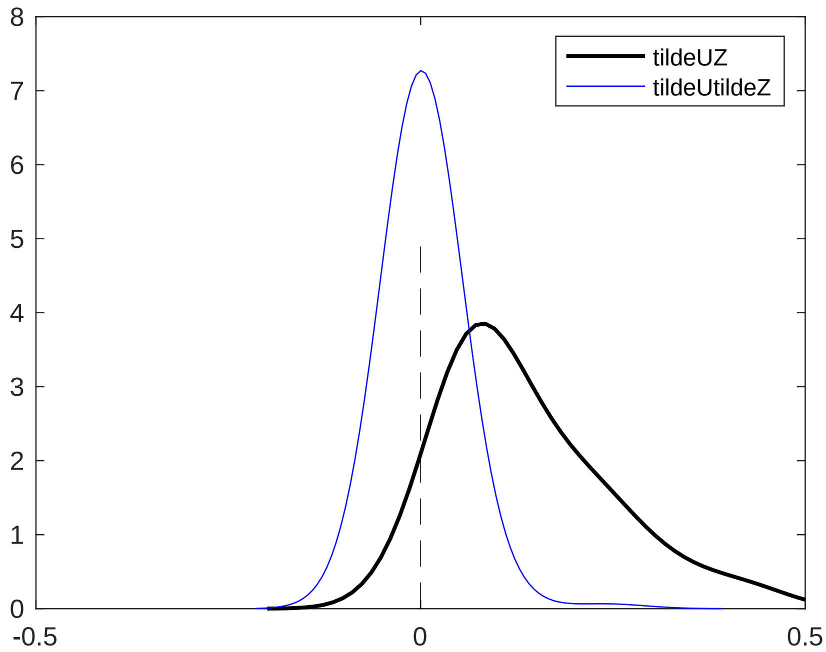
(c) $\rho = \rho_z = \rho_g = 0.5$



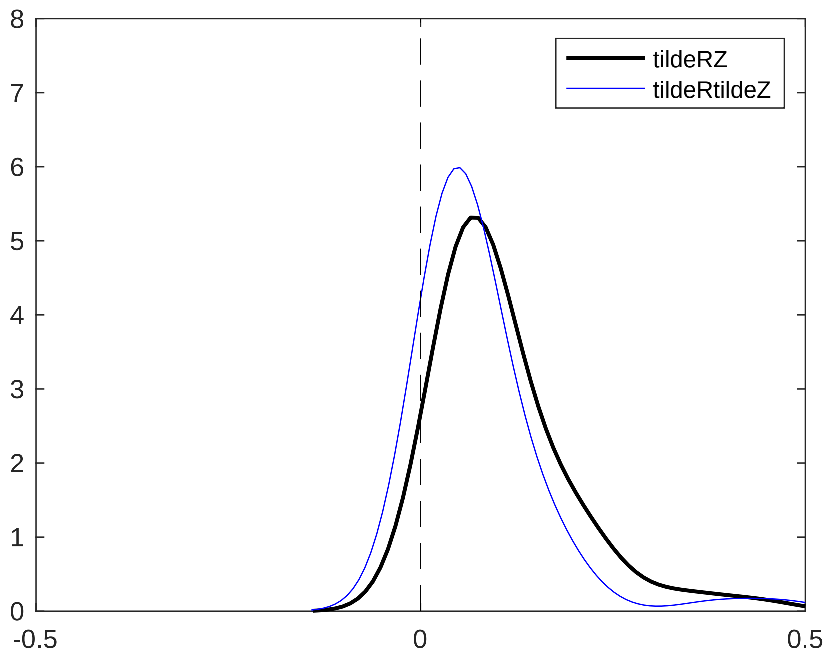
(d) Event study estimates

Note: Panels (a)-(c) display kernel density plots for TWFE OLS, TWFE TSLS and the synthetic IV. Panel (d) shows simulated event study estimates as in Figure 1.2 panel (d) with 95% confidence bands for $\rho = \rho_z = \rho_g = 0.5$. Simulations are done over 1000 iterations with the following parameters: $\beta = \gamma = 1$, $k = 1$, $T = 30$, $T_0 = 20$, $J = 20$, $\sigma_\epsilon = 0.5$, $\kappa = 0.5$, $\sigma_\eta = \sigma_z = \sigma_g = 1$.

Figure A.2: Simulation of finite sample correlations.



(a) Empirical average $\tilde{U}Z$ vs. $\tilde{U}\tilde{Z}$.



(b) Empirical average $\tilde{R}Z$ vs. $\tilde{R}\tilde{Z}$.

Notes: Panel (a) shows the distribution of $\frac{1}{JT}\sum_{it}\tilde{U}_{it}Z_{it}$ and $\frac{1}{JT}\sum_{it}\tilde{U}_{it}\tilde{Z}_{it}$ across 1000 simulations. Panel (b) shows the distribution of $\frac{1}{JT}\sum_{it}\tilde{R}_{it}Z_{it}$ and $\frac{1}{JT}\sum_{it}\tilde{R}_{it}\tilde{Z}_{it}$ across 1000 simulations. The simulation design is that of Figure A.1 with parameters $\rho = \rho_z = \rho_g = 0.5$ as in .

Table A.1: Simulations for different $r = \rho = \rho_z = \rho_g$ and σ_ϵ.

	r=0.5				r=0.7				r=0.9			
	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE
					$\sigma_\epsilon = 0.5$							
OLS (TWFE)	1.31	0.02	0.31	0.11	1.50	0.02	0.50	0.27	1.73	0.01	0.73	0.55
TSLS (TWFE)	1.26	0.07	0.26	0.13	1.51	0.08	0.51	0.34	1.83	0.06	0.83	0.74
SIV	1.02	0.01	0.02	0.01	1.05	0.02	0.05	0.02	1.19	0.04	0.19	0.07
projected SIV	0.92	0.03	-0.08	0.04	0.95	0.04	-0.05	0.05	1.11	0.07	0.11	0.08
Agg. SIV	1.23	0.08	0.23	0.13	1.46	0.08	0.46	0.29	1.80	0.04	0.80	0.68
SIV + projected	1.01	0.01	0.01	0.01	1.03	0.02	0.03	0.02	1.15	0.04	0.15	0.06
SIV + Agg.	1.03	0.01	0.03	0.01	1.07	0.02	0.07	0.02	1.21	0.04	0.21	0.08
SIZ Z	1.07	0.02	0.07	0.02	1.15	0.03	0.15	0.05	1.43	0.03	0.43	0.21
					$\sigma_\epsilon = 1$							
OLS (TWFE)	1.38	0.02	0.38	0.16	1.60	0.02	0.60	0.38	1.86	0.02	0.86	0.76
TSLS (TWFE)	1.26	0.07	0.26	0.14	1.50	0.08	0.50	0.34	1.82	0.06	0.82	0.74
SIV	1.03	0.01	0.03	0.01	1.07	0.03	0.07	0.03	1.26	0.05	0.26	0.12
projected SIV	0.90	0.05	-0.10	0.06	0.94	0.06	-0.06	0.07	1.14	0.08	0.14	0.10
Agg. SIV	1.22	0.07	0.22	0.12	1.47	0.08	0.47	0.30	1.80	0.04	0.80	0.69
SIV + projected	1.01	0.01	0.01	0.01	1.03	0.03	0.03	0.03	1.21	0.05	0.21	0.10
SIV + Agg.	1.04	0.01	0.04	0.02	1.10	0.03	0.10	0.04	1.30	0.05	0.30	0.14
SIZ Z	1.08	0.02	0.08	0.03	1.19	0.03	0.19	0.07	1.50	0.03	0.50	0.28
					$\sigma_\epsilon = 2$							
OLS (TWFE)	1.48	0.02	0.48	0.26	1.74	0.03	0.74	0.58	2.05	0.02	1.05	1.12
TSLS (TWFE)	1.26	0.08	0.26	0.14	1.50	0.09	0.50	0.34	1.82	0.07	0.82	0.74
SIV	1.05	0.03	0.05	0.03	1.12	0.04	0.12	0.06	1.37	0.07	0.37	0.21
projected SIV	0.87	0.08	-0.13	0.09	0.93	0.10	-0.07	0.10	1.21	0.10	0.21	0.15
Agg. SIV	1.22	0.08	0.22	0.12	1.46	0.09	0.46	0.30	1.80	0.05	0.80	0.68
SIV + projected	1.01	0.03	0.01	0.03	1.06	0.05	0.06	0.05	1.29	0.08	0.29	0.16
SIV + Agg.	1.07	0.03	0.07	0.03	1.16	0.05	0.16	0.07	1.43	0.07	0.43	0.26
SIZ Z	1.10	0.03	0.10	0.04	1.24	0.05	0.24	0.10	1.58	0.04	0.58	0.38
					$\sigma_\epsilon = 4$							
OLS (TWFE)	1.63	0.04	0.63	0.43	1.95	0.04	0.95	0.94	2.31	0.04	1.31	1.75
TSLS (TWFE)	1.25	0.09	0.25	0.15	1.49	0.11	0.49	0.35	1.81	0.08	0.81	0.74
SIV	1.08	0.05	0.08	0.05	1.19	0.07	0.19	0.11	1.49	0.10	0.49	0.34
projected SIV	0.85	0.13	-0.15	0.15	0.95	0.15	-0.05	0.15	1.30	0.15	0.30	0.24
Agg. SIV	1.20	0.10	0.20	0.14	1.45	0.11	0.45	0.31	1.79	0.07	0.79	0.69
SIV + projected	1.01	0.05	0.01	0.05	1.10	0.08	0.10	0.09	1.41	0.11	0.41	0.28
SIV + Agg.	1.11	0.05	0.11	0.06	1.24	0.07	0.24	0.13	1.56	0.09	0.56	0.40
SIZ Z	1.13	0.06	0.13	0.07	1.30	0.07	0.30	0.16	1.65	0.06	0.65	0.48
					$\sigma_\epsilon = 8$							
OLS (TWFE)	1.83	0.06	0.83	0.75	2.24	0.08	1.24	1.61	2.68	0.09	1.68	2.91
TSLS (TWFE)	1.24	0.11	0.24	0.17	1.49	0.13	0.49	0.37	1.80	0.11	0.80	0.76
SIV	1.12	0.09	0.12	0.11	1.27	0.12	0.27	0.19	1.61	0.13	0.61	0.50
projected SIV	0.88	0.22	-0.12	0.24	1.01	0.24	0.01	0.24	1.42	0.25	0.42	0.42
Agg. SIV	1.19	0.15	0.19	0.19	1.43	0.15	0.43	0.34	1.78	0.11	0.78	0.72
SIV + projected	1.05	0.10	0.05	0.10	1.18	0.13	0.18	0.17	1.53	0.16	0.53	0.44
SIV + Agg.	1.15	0.10	0.15	0.12	1.32	0.12	0.32	0.23	1.66	0.11	0.66	0.55
SIZ Z	1.16	0.09	0.16	0.12	1.36	0.11	0.36	0.24	1.71	0.09	0.71	0.59

method. Imbens and Viviano (2023) highlight a similar comparison and note that: "*Two main differences from properties of least squares estimators as in Bai (2009) is that Synthetic Control does not require a (i) low-rank assumption on the factors, but only a low-rank assumption on the*

Table A.2: $T_0 = 20,\ J = 20,\ \sigma_\epsilon = 0.5,\ \sigma_z = 1,\ \sigma_{other} = 0.5,\ \kappa = 0.5.$

Coverage $\alpha = 0.05$			
	T=30	T=40	T=50
$\rho = \rho_g = \rho_z = 0.0$	0.981	0.962	0.952
$\rho = \rho_g = \rho_z = 0.3$	0.976	0.944	0.96
$\rho = \rho_g = \rho_z = 0.5$	0.960	0.945	0.923
$\rho = \rho_g = \rho_z = 0.7$	0.904	0.808	0.792

endogenous factors; (ii) it does not require to specify or estimate the number of (endogenous) factors." Intuitively, there might be advantages in using SCs, as the method provides useful regularization through the l_1-norm or simplex constraint, and is less dependent on an exact low-rank representation. Furthermore, relative to Bai (2009) which require both the number of periods and units to be large, our consistency result requires that T_0 and JT are large, meaning that our method is valid for either fixed J or fixed T. These rate differences and regularization advantages are likely to translate into better finite sample performance, especially in settings with moderate rank.

It is also an open question what are the statistical properties of using plug-in factor model estimates in our setting. Panel data approaches for treatment effect estimation with factor model plug-ins, such as the one described by Li and Sonnier (2023), rely on the existence of a donor pool of never-treated units for the consistent estimation of the common factors F_t in the post-treatment period ($t > T_0$). In our setting, because all units are treated in the post-treatment period an alternative procedure needs to be used. We propose the following procedure that exploits the pre-treatment period to consistently estimate the factor loadings using PCA/SVD and jointly estimate the common factor and θ_t parameters in an iterative procedure as proposed in Bai (2009).

1. For $t \leq T_0$ we estimate μ_i by truncated SVD: Let $Y^{T_0\prime}$ denote the $J \times T$ matrix of pre-treatment outcomes for the case in which $R_{it} = 0$ for $t \leq T_0$ such that $Y_{it} = \mu_i' F_t + \epsilon_{it}$. Then, we estimate the factor loadings as $\hat{\mu} = \hat{U}^k \hat{S}^k$ for estimates of the $k-$truncated SVD $Y^{T_0\prime} = USV'$.

2. For $t > T_0$ we estimate F and θ jointly by iteratively solving for θ by ordinary least squares and for F by PCA. We iterate until convergence the following two equations

$$\hat{\theta}_{OLS}(F) = (R' M_{\hat{F}} R)^{-1} R' M_{\hat{F}} Y,$$

   where M_F is the residual maker matrix for the estimated common factors $\hat{F}$, and from the estimates of the k-truncated SVD of $Y = Y^{T_0\prime} - \hat{\theta} R^{T_0\prime}$, the common factors are given by $\hat{F} = \hat{V}^k$.

3. Given the estimated $\hat{\mu}_i' \hat{F}_t$ we compute the IV-PCA estimator $\hat{\theta}_{IV-PCA}$ as the standard TSLS with covariates (i.e. with $\hat{\mu}_i' \hat{F}_t$ projected out).

We compare the different methods under the simulation design calibrated to the Syrian example

Table A.3: Simulations for $T_0 = 10$

	r=0.5				r=0.7				r=0.9			
	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE
					$\sigma = 0.5$							
OLS (TWFE)	1.29	0.02	0.29	0.10	1.48	0.02	0.48	0.25	1.71	0.01	0.71	0.52
TSLS (TWFE)	1.23	0.05	0.23	0.11	1.46	0.07	0.46	0.29	1.78	0.06	0.78	0.68
SIV	1.01	0.01	0.01	0.01	1.07	0.02	0.07	0.02	1.25	0.04	0.25	0.10
projected SIV	0.93	0.04	-0.07	0.04	0.99	0.05	-0.01	0.05	1.24	0.36	0.24	0.41
Agg. SIV	1.23	0.06	0.23	0.12	1.47	0.07	0.47	0.30	1.79	0.05	0.79	0.68
SIV + projected	0.94	0.04	-0.06	0.04	0.99	0.05	-0.01	0.05	1.24	0.35	0.24	0.41
SIV + Agg.	1.23	0.06	0.23	0.12	1.47	0.07	0.47	0.29	1.79	0.05	0.79	0.67
SIZ Z	1.05	0.02	0.05	0.02	1.16	0.03	0.16	0.05	1.49	0.03	0.49	0.27
					$\sigma = 1$							
OLS (TWFE)	1.36	0.02	0.36	0.15	1.59	0.02	0.59	0.36	1.84	0.01	0.84	0.73
TSLS (TWFE)	1.22	0.06	0.22	0.11	1.46	0.07	0.46	0.29	1.78	0.06	0.78	0.67
SIV	1.02	0.02	0.02	0.02	1.11	0.03	0.11	0.04	1.35	0.05	0.35	0.17
projected SIV	0.93	0.05	-0.07	0.06	1.00	0.07	0.00	0.07	0.66	44.22	-0.34	44.29
Agg. SIV	1.24	0.06	0.24	0.12	1.47	0.07	0.47	0.29	1.80	0.05	0.80	0.69
SIV + projected	0.93	0.05	-0.07	0.06	1.00	0.07	0.00	0.07	0.67	43.34	-0.33	43.40
SIV + Agg.	1.24	0.06	0.24	0.12	1.47	0.07	0.47	0.29	1.79	0.05	0.79	0.68
SIZ Z	1.06	0.02	0.06	0.03	1.21	0.03	0.21	0.07	1.56	0.03	0.56	0.34
					$\sigma = 2$							
OLS (TWFE)	1.47	0.02	0.47	0.24	1.73	0.02	0.73	0.56	2.03	0.02	1.03	1.08
TSLS (TWFE)	1.22	0.06	0.22	0.11	1.46	0.08	0.46	0.29	1.78	0.06	0.78	0.67
SIV	1.04	0.03	0.04	0.03	1.17	0.05	0.17	0.07	1.47	0.06	0.47	0.28
projected SIV	0.93	0.08	-0.07	0.08	1.03	0.10	0.03	0.10	1.34	0.13	0.34	0.25
Agg. SIV	1.25	0.08	0.25	0.14	1.47	0.09	0.47	0.31	1.80	0.05	0.80	0.70
SIV + projected	0.93	0.08	-0.07	0.08	1.04	0.09	0.04	0.10	1.34	0.13	0.34	0.25
SIV + Agg.	1.24	0.07	0.24	0.13	1.47	0.09	0.47	0.31	1.80	0.05	0.80	0.69
SIZ Z	1.08	0.03	0.08	0.04	1.26	0.04	0.26	0.11	1.62	0.04	0.62	0.43
					$\sigma = 4$							
OLS (TWFE)	1.62	0.04	0.62	0.42	1.94	0.04	0.94	0.92	2.30	0.03	1.30	1.72
TSLS (TWFE)	1.21	0.08	0.21	0.12	1.45	0.09	0.45	0.29	1.77	0.07	0.77	0.67
SIV	1.06	0.05	0.06	0.06	1.23	0.07	0.23	0.12	1.58	0.08	0.58	0.42
projected SIV	0.94	0.13	-0.06	0.14	1.08	0.15	0.08	0.15	1.46	0.15	0.46	0.35
Agg. SIV	1.24	0.09	0.24	0.15	1.46	0.12	0.46	0.34	1.80	0.06	0.80	0.70
SIV + projected	0.95	0.13	-0.05	0.13	1.08	0.15	0.08	0.15	1.46	0.15	0.46	0.35
SIV + Agg.	1.24	0.09	0.24	0.15	1.46	0.12	0.46	0.33	1.80	0.06	0.80	0.70
SIZ Z	1.11	0.05	0.11	0.06	1.32	0.05	0.32	0.16	1.68	0.05	0.68	0.51
					$\sigma = 8$							
OLS (TWFE)	1.82	0.06	0.82	0.74	2.23	0.07	1.23	1.58	2.67	0.07	1.67	2.87
TSLS (TWFE)	1.20	0.11	0.20	0.15	1.44	0.12	0.44	0.31	1.77	0.09	0.77	0.68
SIV	1.08	0.09	0.08	0.10	1.29	0.11	0.29	0.19	1.66	0.11	0.66	0.55
projected SIV	0.96	0.20	-0.04	0.20	1.15	0.22	0.15	0.24	1.55	0.23	0.55	0.52
Agg. SIV	1.22	0.13	0.22	0.18	1.46	0.16	0.46	0.37	1.79	0.09	0.79	0.72
SIV + projected	0.97	0.19	-0.03	0.19	1.15	0.22	0.15	0.24	1.55	0.22	0.55	0.52
SIV + Agg.	1.22	0.13	0.22	0.17	1.46	0.16	0.46	0.37	1.79	0.09	0.79	0.72
SIZ Z	1.13	0.07	0.13	0.09	1.36	0.08	0.36	0.21	1.73	0.06	0.73	0.60

as in Table 1.1. While in the main text we considered different noise levels, here we focus on designs with different numbers of factors k and different correlations between the instrument Z_{it} and the factor structure $\mu_i' F_t$. Table A.4 displays the results of the simulations. We find that the IV-PCA estimator performs well in the low-rank case $k = 1$, in which it has a similar MSE to the SIV, while exhibiting a somewhat larger bias. However, as we increase the number of factors to $k = 2, 3$ and 5, we find that IV-PCA performs significantly worse. The SIV exhibits negligible finite sample bias in medium correlation settings, and even in the $k = 5$ case, the estimator is significantly less biased than the TSLS, this is not true for the IV-PCA. As we increase k the estimator becomes more and more biased, reaching similar levels to the TSLS for $k = 5$.

These results highlight the same phenomenon shown in the simulations of Imbens and Viviano (2023), in which the PCA based estimator has significantly larger finite sample bias than the SC estimator when the number of endogenous factors is large (in our case all of the factors are endogenous).

Table A.4: Simulations comparing SIV and IV-PCA. Design calibrated to the Syrian example, for different correlations $\rho = \rho_z = \rho_g = r$ and number of factors k.

	r=0.5				r=0.7				r=0.9			
	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE	Mean	Var	Bias	MSE
					k=1							
OLS (TWFE)	0.007	0.018	0.167	0.046	0.125	0.025	0.285	0.106	0.274	0.023	0.434	0.211
TSLS (TWFE)	-0.051	0.026	0.109	0.038	0.053	0.036	0.213	0.081	0.193	0.031	0.353	0.155
SIV	-0.151	0.005	0.009	0.005	-0.134	0.007	0.026	0.008	-0.062	0.016	0.098	0.025
IV-PCA	-0.140	0.003	0.020	0.003	-0.113	0.005	0.047	0.007	-0.012	0.021	0.148	0.043
					k=2							
OLS (TWFE)	-0.028	0.011	0.132	0.028	0.084	0.014	0.244	0.073	0.230	0.012	0.390	0.164
TSLS (TWFE)	-0.053	0.012	0.107	0.023	0.053	0.015	0.213	0.061	0.196	0.013	0.356	0.139
SIV	-0.144	0.002	0.016	0.002	-0.116	0.003	0.044	0.005	-0.019	0.006	0.141	0.026
IV-PCA	-0.081	0.008	0.079	0.014	0.016	0.012	0.176	0.043	0.168	0.010	0.328	0.118
					k=3							
OLS (TWFE)	-0.043	0.008	0.117	0.022	0.065	0.010	0.225	0.061	0.210	0.008	0.370	0.145
TSLS (TWFE)	-0.059	0.009	0.101	0.019	0.046	0.011	0.206	0.053	0.188	0.009	0.348	0.130
SIV	-0.141	0.002	0.019	0.002	-0.105	0.002	0.055	0.006	0.017	0.005	0.177	0.036
IV-PCA	-0.071	0.007	0.089	0.015	0.031	0.008	0.191	0.044	0.190	0.006	0.350	0.128
					k=5							
OLS (TWFE)	-0.044	0.006	0.116	0.019	0.065	0.007	0.225	0.057	0.209	0.005	0.369	0.141
TSLS (TWFE)	-0.052	0.006	0.108	0.017	0.054	0.007	0.214	0.053	0.196	0.006	0.356	0.133
SIV	-0.124	0.002	0.036	0.003	-0.070	0.003	0.090	0.011	0.075	0.004	0.235	0.059
IV-PCA	-0.057	0.005	0.103	0.015	0.047	0.006	0.207	0.048	0.197	0.003	0.357	0.131

A.2 Data

Turkish Statistical Institute (Turkstat) defines employment under four categories: wage-employment (60.7%), self-employment (20.3%), unpaid family worker (13.2%) and employer (5.6%). Wage-employment, or salaried employment, refers to the type of jobs that are done as an exchange for monetary or non-monetary payment. Both fixed and hourly pay are considered wage-employment under this category. The reason why we focus on salaried employment as opposed to overall employment for the empirical section of the paper is that, as suggested by Gulek (2025), wage employment and non-wage employment (self-employment, employer, or unpaid family work) are driven by different economic forces. Whereas there has to be an employer willing to hire a worker for a particular wage for that worker to have a salaried job (i.e, we can think about a labor demand curve), self-employment is an individual labor-supply decision. Natives who lose their salaried jobs due to the labor supply shock may choose to search for a salaried job while remaining unemployed, or if self-employment is a feasible alternative, may choose to remain employed. Gulek (2025) shows that transition from salaried to non-salaried jobs is an important adjustment mechanism for Turkish men but not so for Turkish women. Whereas he finds similar effects for men and women in salaried employment, he finds opposing results for non-salaried employment. He further argues that the canonical labor demand framework is more appropriate to think about wage employment (as opposed to non-wage employment) in settings where self-employment is a feasible alternative.

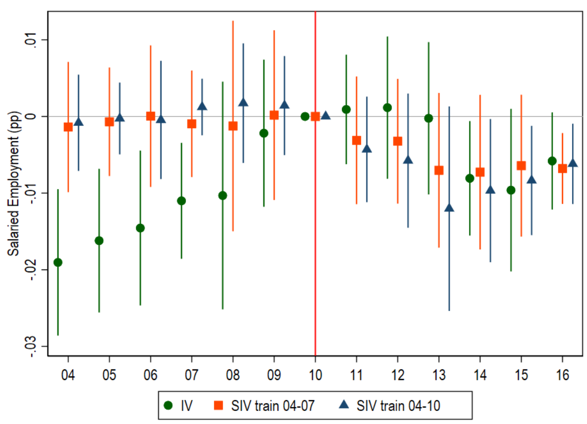
Table A.5: Educational Attainment of Syrian refugees and Natives

Educational Attainment	Syrian migrants (age 18+)	Natives (Age: 18-64)
No degree	0.21	0.12
Primary school	0.42	0.33
Secondary school	0.20	0.16
High school	0.10	0.20
Some college and above	0.08	0.19

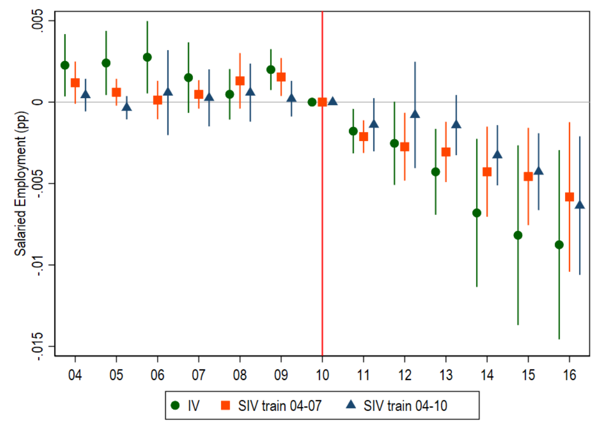
Source: Author's calculation using 2019 Household Labor Force Survey for natives, and Turkish Red Crescent and WFP (2019) for the Syrian refugees.

In the main text, we write that Syrian refugees are less educated than the Turkish natives. We show evidence for this on Table A.5. We use Turkish Household Labor force Surveys to determine the educational attainment of natives, and use livelihood surveys that are conducted on Syrian refugees to determine their educational attainment. According to these surveys, 21% of Syrian refugees in Turkey do not have any degree, 63% have at most a primary school degree, and 83% do not have a high school diploma, whereas these numbers are 12%, 45%, and 61%, respectively for natives.

Figure A.3: Additional examples of IV vs SIV



(a) Men: salaried employment



(b) Women: formal salaried employment

Notes: This Figure replicates Figure 1.4 with an additional panel for formal women salaried employment.

A.3 Replication of Autor et al. (2013)

Table A.6: Replication of Table 3 in Autor et al. (2013)

	1990–2007 stacked first differences					
	(1)	(2)	(3)	(4)	(5)	(6)
IV	-0.75	-0.61	-0.54	-0.51	-0.56	-0.60
	(0.07)	(0.09)	(0.09)	(0.08)	(0.10)	(0.10)
SIV	-0.70	-0.59	-0.51	-0.50	-0.61	-0.63
	(0.07)	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)
Controls						
Percentage of employment in manufacturing t-1	No	Yes	Yes	Yes	Yes	Yes
Percentage of college-educated population t-1	No	No	No	Yes	No	Yes
Percentage of foreign-born population t-1	No	No	No	Yes	No	Yes
Percentage of employment among women t-1	No	No	No	Yes	No	Yes
Percentage of employment in routine occupations t-1	No	No	No	No	Yes	Yes
Average offshorability index of occupations t-1	No	No	No	No	Yes	Yes
Census division dummies	No	No	Yes	Yes	Yes	Yes

Notes: The first row replicates columns 1–6 of Table 3 in ADH 2013. Row 2 shows SIV estimates. The SC weights are estimated using the manufacturing growth rates in 1970 and 1980. Dependent variable: 10 $\times$ annual change in manufacturing emp/working-age pop (in % pts). N = 1,444 (722 commuting zones $\times$ 2 time periods). All regressions include a constant and a dummy for the 2000–2007 period. Routine occupations are defined such that they account for 1/3 of US employment in 1980. The offshorability index variable is standardized to mean of 0 and standard deviation of 10 in 1980. Robust standard errors in parentheses are clustered on state. Models are weighted by start of period CZ share of national population.

A.4 Additional figure for rank effects

Figure A.4: Reduced-form estimates using the 1990 and 2000 shares

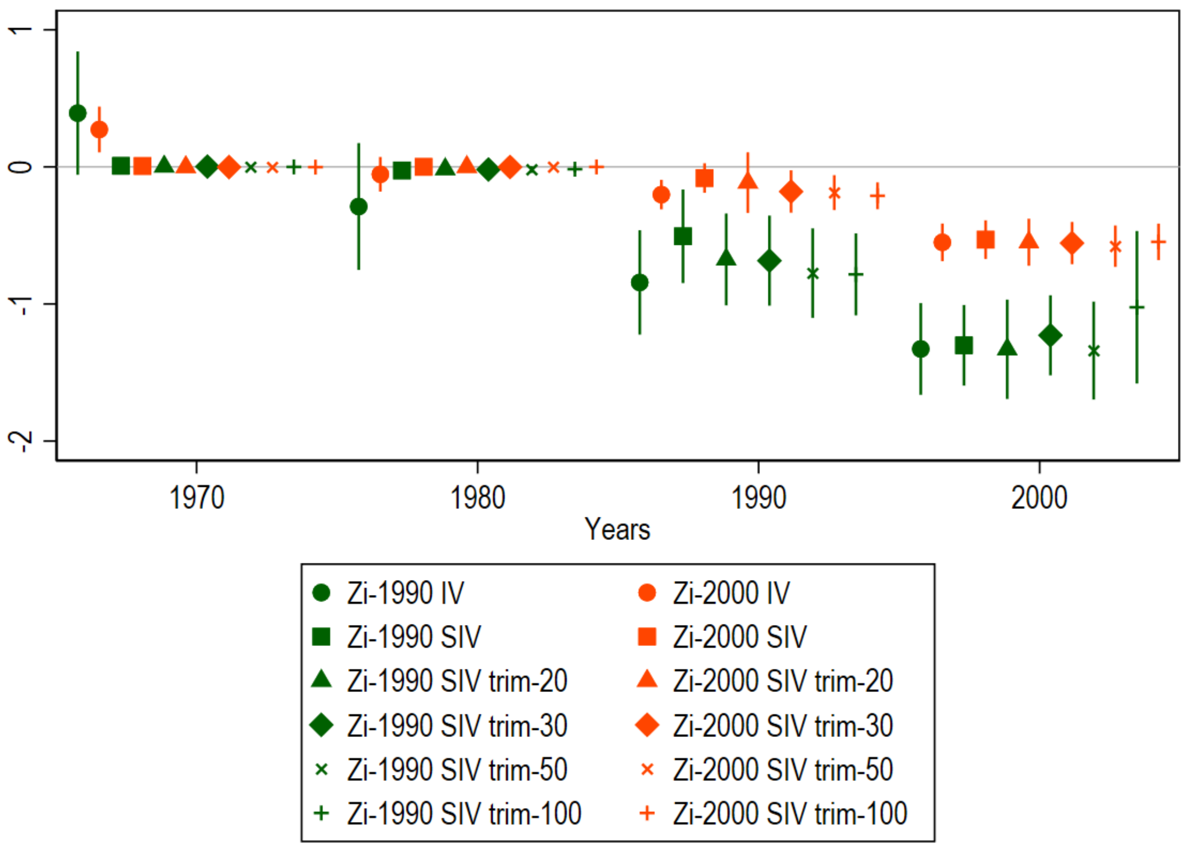


Table A.7: Replication of Table 2 in Autor et al. (2013)

	1990–2000	2000-2007	1990–2007
	(1)	(2)	(3)
2SLS	-0.888	-0.718	-0.746
	(0.181)	(0.064)	(0.068)
SIV	-0.588	-0.726	-0.703
	(0.198)	(0.070)	(0.067)
SIV-trim20	-0.752	-0.763	-0.761
	(0.178)	(0.104)	(0.096)
SIV-trim30	-0.784	-0.769	-0.772
	(0.177)	(0.102)	(0.089)
SIV-trim50	-0.874	-0.807	-0.819
	(0.172)	(0.081)	(0.078)
SIV-trim100	-0.937	-0.769	-0.801
	(0.170)	(0.067)	(0.069)

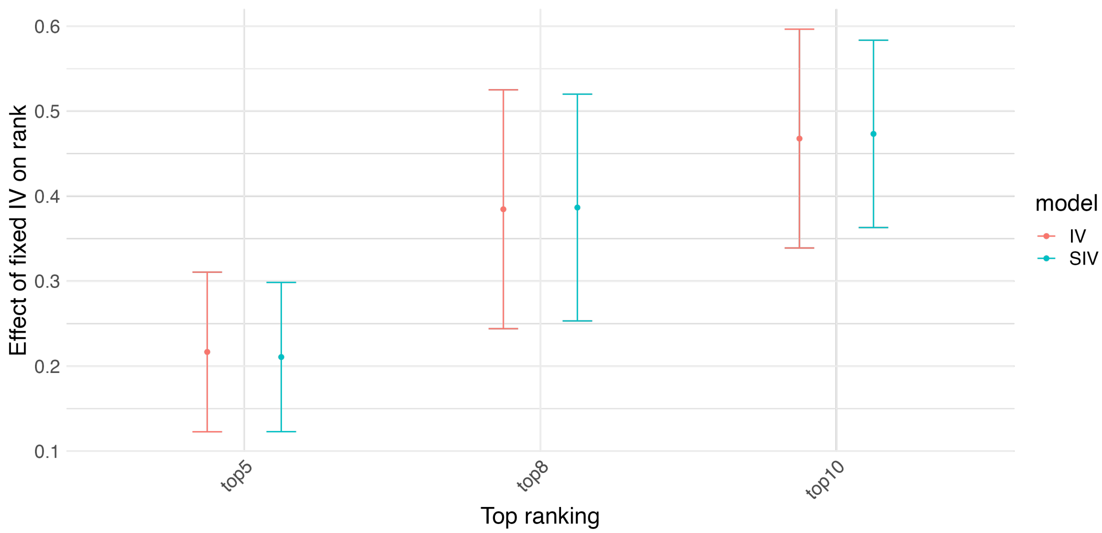
Notes: The first row replicates columns 1–3 of Table 2 in ADH 2013. In rows 2–6, we apply SIV. The SC weights are estimated using the manufacturing growth rates in 1970 and 1980. Rows 3, 4, 5, and 6 show the SIV with the donor pool trimmed to the 20, 30, 50, an 100 closest closest units to the treated unit according to the Euclidean distance, respectively.

Table A.8: Replication of Table 3 in Autor et al. (2013) with trimming

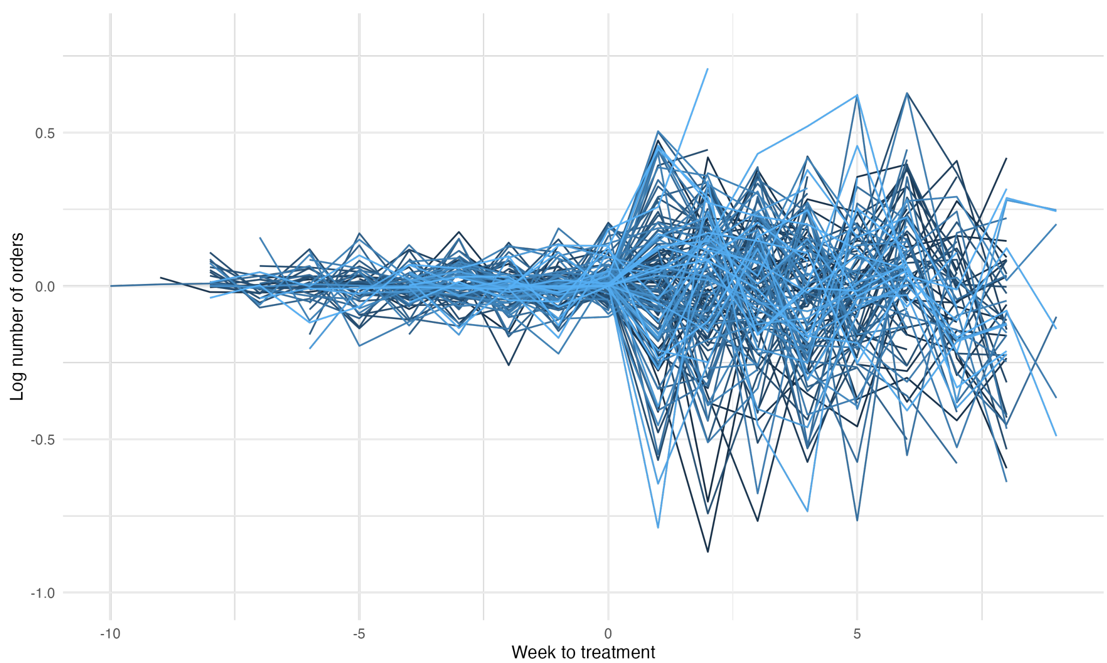
	1990–2007 stacked first differences					
	(1)	(2)	(3)	(4)	(5)	(6)
IV	-0.75	-0.61	-0.54	-0.51	-0.56	-0.60
	(0.07)	(0.09)	(0.09)	(0.08)	(0.10)	(0.10)
SIV	-0.70	-0.59	-0.51	-0.50	-0.61	-0.63
	(0.07)	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)
SIV-trim20	-0.76	-0.67	-0.58	-0.57	-0.65	-0.65
	(0.10)	(0.10)	(0.07)	(0.08)	(0.08)	(0.07)
SIV-trim30	-0.77	-0.66	-0.54	-0.53	-0.59	-0.61
	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)
SIV-trim50	-0.82	-0.74	-0.63	-0.62	-0.68	-0.70
	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.09)
SIV-trim100	-0.80	-0.73	-0.63	-0.62	-0.67	-0.69
	(0.07)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
Controls						
Percentage of employment in manufacturing t-1	No	Yes	Yes	Yes	Yes	Yes
Percentage of college-educated population t-1	No	No	No	Yes	No	Yes
Percentage of foreign-born population t-1	No	No	No	Yes	No	Yes
Percentage of employment among women t-1	No	No	No	Yes	No	Yes
Percentage of employment in routine occupations t-1	No	No	No	No	Yes	Yes
Average offshorability index of occupations t-1	No	No	No	No	Yes	Yes
Census division dummies	No	No	Yes	Yes	Yes	Yes

Notes: The first row replicates columns 1–6 of Table 3 in ADH 2013. Row 2 shows SIV estimates. The SC weights are estimated using the manufacturing growth rates in 1970 and 1980. Dependent variable: 10 $\times$ annual change in manufacturing emp/working-age pop (in % pts). N = 1,444 (722 commuting zones $\times$ 2 time periods). All regressions include a constant and a dummy for the 2000–2007 period. Routine occupations are defined such that they account for 1/3 of US employment in 1980. The offshorability index variable is standardized to mean of 0 and standard deviation of 10 in 1980. Robust standard errors in parentheses are clustered on state. Models are weighted by start of period CZ share of national population. Rows 3, 4, 5, and 6 show the SIV with the donor pool trimmed to the 20, 30, 50, an 100 closest closest units to the treated unit according to the Euclidean distance, respectively.

Figure A.5: First stage and SIV fit.



(a) First stage



(b) $\tilde{Y}_{it}$ fit.

Notes: Panel (a) shows the first stage regression estimates of log number of orders on R_{it}^k with week and store fixed effects. Panel (b) plots $\tilde{Y}_{it}$ (debiased log number of orders) for each producer, where we keep the 70 producers with best fit.

Appendix B

Appendix to Chapter 2

B.1 Conditional distribution model derivations

Suppose we have the following simple model independent of time:

$$Y_i = \lambda_i F_t + \epsilon_i,$$

where $F_t \sim N(0, \sigma^2)$ and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Then it follows that $\mathbb{E}[Y_1] = 0$, $\mathbb{E}[\boldsymbol{Y}_J] = \boldsymbol{0}$, $cov(Y_i, Y_j) = \lambda_i \lambda_j \sigma^2$ and $var(Y_j) = \sigma_\epsilon^2 + \lambda_i^2 \sigma^2$. Then, the joint distribution of ys is normal. We are interested in the conditional distribution of Y_1 given $\boldsymbol{y}_J = (y_2, \dots, y_{J+1})$:

$$Y_1 | \boldsymbol{y}_J \sim N\left(\tilde{\mu}, \tilde{\Sigma}\right),$$

where

$$\tilde{\mu} = cov(Y_1, \boldsymbol{Y}_J)\Sigma_{(2,J+1)}^{-1}\boldsymbol{y}_J = \sum_{j=2}^{J+1} w_j(\boldsymbol{\lambda}, \sigma) y_j \tag{B.1}$$

$$\tilde{\Sigma} = var(Y_1) - cov(Y_1, \boldsymbol{Y}_J)\Sigma_{(2,J+1)}^{-1} cov(\boldsymbol{Y}_J, Y_1). \tag{B.2}$$

In this set up the data are repeated observations over time. For a pre-treatment period of length T_0 then we will be interested in estimators $\hat{w}_j$ such that our estimated counterfactual is given by:

$$\hat{Y}_{1t}(0) = \sum_{j=2}^{J+1} \hat{w}_j y_{jt}.$$

Start by noting that in this setting $\Sigma_{(2,J+1)}$ is positive definite and invertible. Hence, by the spectral theorem we can express it as a linear combination of eigenvalues and eigenvectors:

$$\Sigma_{(2,J+1)} = \begin{pmatrix} \sigma_\epsilon^2 + \lambda_2^2\sigma^2 & \lambda_2\lambda_3\sigma^2 & \cdots & \lambda_2\lambda_{J+1}\sigma^2 \\ \lambda_2\lambda_3\sigma^2 & \sigma_\epsilon^2 + \lambda_3^2\sigma^2 & \cdots & \lambda_3\lambda_{J+1}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_2\lambda_{J+1}\sigma^2 & \lambda_3\lambda_{J+1}\sigma^2 & \cdots & \sigma_\epsilon^2 + \lambda_{J+1}^2\sigma^2 \end{pmatrix} = \sum_{j=2}^{J+1} s_j \boldsymbol{u}_j \boldsymbol{u}_j^T,$$

where s_j is the eigenvalue associated with the $\boldsymbol{u}_j$ eigenvector. Observe that the eigenvalues are given by $s_2 = \cdots = s_J = 1$ and $s_{J+1} = \sigma_\epsilon^2 + \sum_{j=2}^{J+1} \lambda_j^2 \sigma^2$. Therefore, we can express $\tilde{\mu}$ as a linear combination of the data, $\boldsymbol{\lambda}$ and σ:

$$\tilde{\mu} = \sigma^2\lambda_1 \sum_{j=2}^{J+1}\sum_{i=2}^{J+1} \lambda_i [\Sigma^{-1}_{(2,J+1)}]_{ji} y_j$$
$$= \sum_{j=2}^{J+1} \sigma^2\lambda_1 \sum_{i=2}^{J+1} \lambda_i \sum_{k=2}^{J+1} \frac{1}{s_k} [\boldsymbol{u}_k \boldsymbol{u}_k^T]_{ji} y_j.$$

By substituting the eigenvectors and eigenvalues we can find the closed form for the weights:

$$w_j(\boldsymbol{\lambda}, \sigma) = \sigma^2\lambda_1 \sum_{i=2}^{J+1} \lambda_i \sum_{k=2}^{J+1} \frac{1}{s_k} [\boldsymbol{u}_k \boldsymbol{u}_k^T]_{ji}$$
$$= \frac{\sigma^2\lambda_1\lambda_j}{\sigma_\epsilon^2 + \sum_{j=2}^{J+1} \lambda_j^2 \sigma^2}.$$

Similarly, we can derive a closed form for the variance:

$$\tilde{\Sigma} = \sigma_\epsilon^2 + \lambda_1\sigma^2 - \frac{\sigma^4\lambda_1^2 \sum_{j=2}^{J+1} \lambda_j^2}{\sigma_\epsilon^2 + \sigma^2 \sum_{j=2}^{J+1} \lambda_j^2}$$
$$= \sigma_\epsilon^2 + \lambda_1\sigma^2(\sigma_\epsilon^2 - \sum_{j=2}^{J+1} w_j\lambda_j)$$
$$= \sigma_\epsilon^2 + \lambda_1\sigma^2 - (\sigma_\epsilon^2 + \sigma^2 \sum_{j=2}^{J+1} \lambda_j^2) \sum_{j+2}^{J+1} w_j^2.$$

B.2 Proof of Theorem 1

Observe that without loss of generality we can consider $V = I$. For notational convenience we drop the potential outcome subscript in what follows such that $\boldsymbol{Y}_1 \equiv \boldsymbol{Y}_1(0)$. Observe that this proof is equivalent to showing that the conditional expectation is the best linear predictor under the square loss. Consider the objective function:

$$\mathbb{E}\left[(\boldsymbol{Y}_1 - \boldsymbol{y}_J'\boldsymbol{w})'(\boldsymbol{Y}_1 - \boldsymbol{y}_J'\boldsymbol{w})\right] = \mathbb{E}\left[(\boldsymbol{Y}_1'\boldsymbol{Y}_1 - 2\boldsymbol{w}'\boldsymbol{y}_J\boldsymbol{Y}_1 + \boldsymbol{w}'\boldsymbol{y}_J\boldsymbol{y}_J'\boldsymbol{w}\right]$$
$$= \mathbb{E}\left[\mathbb{E}[\boldsymbol{Y}_1'\boldsymbol{Y}_1|\boldsymbol{Y}_J] - 2\boldsymbol{w}'\boldsymbol{y}_J\mathbb{E}[\boldsymbol{Y}_1|\boldsymbol{Y}_J] + \boldsymbol{w}'\boldsymbol{y}_J\boldsymbol{y}_J'\boldsymbol{w}\right]$$
$$= \mathbb{E}\left[\tilde{\boldsymbol{w}}'\boldsymbol{Y}_J\boldsymbol{Y}_J'\tilde{\boldsymbol{w}} + T_0\tilde{\Sigma}^2 - 2\boldsymbol{w}'\boldsymbol{y}_J\boldsymbol{Y}_J'\tilde{\boldsymbol{w}} + \boldsymbol{w}'\boldsymbol{y}_J\boldsymbol{y}_J'\boldsymbol{w}\right]$$

where the steps follow from the derivations for the conditional multivariate normal, the law of iterated expectations and the definition of conditional variance, $V(Y|X) + \mathbb{E}[Y|X]^2 = E[Y^2|X]$. Given that the w parameters are only on the observed $\boldsymbol{y}$, we only can restrict attention to the event $\{\boldsymbol{Y}_J = \boldsymbol{y_j}\}$. Conditional on this event, it is straightforward to see that setting $\boldsymbol{w} = \tilde{\boldsymbol{w}}$ minimizes the expression. Therefore, it follows that conditional on the data, $\tilde{\boldsymbol{w}}$ is a global minimizer of the expression.

B.3 Proof of Theorem 2

Statement (1): we want to show that as $J \to \infty$

$$\mathbb{E}\left[(\boldsymbol{y}'_{JT_0+1}\tilde{\boldsymbol{w}} - \lambda_1 F_{T_0+1})^2\right] \to 0.$$

Expanding the expression conditional on F_{T_0+1} and given the assumptions and derivations for $\tilde{\boldsymbol{w}}$:

$$\begin{aligned}
\mathbb{E}\left[(\boldsymbol{y}'_{JT_0+1}\tilde{\boldsymbol{w}} - \lambda_1 F_{T_0+1})^2 | F_{T_0+1}\right] &= \frac{\lambda_1^2\sigma^4}{(\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2)^2}\mathbb{E}\left[\left(F_{T_0+1}\|\boldsymbol{\lambda}_J\|_2^2 + \sum_j \lambda_j^2\epsilon_{jT_0+1}\right)^2 |F_{T_0+1}\right] \\
&+ \lambda_1^2 F_{T_0+1}^2\left(\frac{\sigma_\epsilon^2 - \sigma^2\|\boldsymbol{\lambda}_J\|_2^2}{\sigma_\epsilon^2 + \sigma^2\|\boldsymbol{\lambda}_J\|_2^2}\right) \\
&- \frac{2\sigma^2\lambda_1^2}{\sigma_\epsilon^2 + \sigma^2\|\boldsymbol{\lambda}_J\|_2^2}\mathbb{E}\left[\sum_j \lambda_j\epsilon_{jT_0+1}|F_{T_0+1}\right] \\
&= \frac{\lambda_1^2\sigma^4}{(\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2)^2}\left(F_{T_0+1}^2\|\boldsymbol{\lambda}_J\|_2^4 + \|\boldsymbol{\lambda}_J^2\|_2^2\right) \\
&+ \lambda_1^2 F_{T_0+1}^2\left(\frac{\sigma_\epsilon^2 - \sigma^2\|\boldsymbol{\lambda}_J\|_2^2}{\sigma_\epsilon^2 + \sigma^2\|\boldsymbol{\lambda}_J\|_2^2}\right)
\end{aligned}$$

Observe that there exists no value of $\|\boldsymbol{\lambda}_J\|_2^2$ for which the above expression is zero. By the law of iterated expectations this implies that convergence in mean squared can not be achieved unless $\sigma_\epsilon^2 = 0$.

Statement (2): By Markov's inequality we have that a for $t > 0$

$$\mathbb{P}(|\boldsymbol{y}'_{JT_0+1}\tilde{\boldsymbol{w}} - \lambda_1 F_{T_0+1}| \geq t) \leq \frac{\mathbb{E}[|\boldsymbol{y}'_{JT_0+1}\tilde{\boldsymbol{w}} - \lambda_1 F_{T_0+1}|]}{t}.$$

Under the assumption that independently $\epsilon_{jt} \sim N(0, \sigma_\epsilon^2)$:

$$\begin{aligned}
\mathbb{E}\left[\left|\boldsymbol{y}'_{JT_0+1}\tilde{\boldsymbol{w}} - \lambda_1 F_{T_0+1}\right| |F_{T_0+1}\right] &= \mathbb{E}\left[\left|\frac{\sigma^2\lambda_1}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2}\left(\|\boldsymbol{\lambda}_J\|_2^2 F_{T_0+1} + \sum_j \lambda_j \epsilon_{jT_0+1}\right) - \lambda_1 F_{T_0+1}\right| |F_{T_0+1}\right] \\
&= \mathbb{E}\left[\left|\frac{-F_{T_0+1}\sigma^2\lambda_1}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2} + \frac{\sigma^2\lambda_1}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2}\sum_j \lambda_j \epsilon_{jT_0+1}\right| |F_{T_0+1}\right] \\
&\leq \left|\frac{\sigma^2\lambda_1 F_{T_0+1}}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2}\right| + \left|\frac{\sigma^2\lambda_1}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2}\right| \mathbb{E}\left[\left|\sum_j \lambda_j \epsilon_{jT_0+1}\right| |F_{T_0+1}\right] \\
&\leq \left|\frac{\sigma^2\lambda_1 F_{T_0+1}}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2}\right| + \left|\frac{\sigma^2\lambda_1}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2}\right| \sum_j |\lambda_j| \frac{\sqrt{2}\sigma_\epsilon}{\sqrt{\pi}}.
\end{aligned}$$

The last step uses the mean of the half-normal distribution. Hence, by the law of iterated expectations convergence in probability holds if as $J \to \infty$, $\frac{1}{\|\boldsymbol{\lambda}_J\|_2^2}\sum_j |\lambda_j| \to 0$. Observe, that by Statement 1 we know that this proof would not work with the application of Chebyshev's inequality.

B.4 Proof of Theorem 3

The first part of the theorem follows directly from the derivation for $\tilde{\boldsymbol{w}}$ given A1-A2, setting

$$\frac{\sigma^2\lambda_1\sum_{j=2}^{J+1}\lambda_j}{\sigma_\epsilon^2 + \sum_{j=2}^{J+1}\lambda_j^2\sigma^2} = 1,$$

and simplifying yields the second condition. For the sign restriction, given that the denominator of $\tilde{w}_j$ is non-negative,

$$\tilde{w}_j = \frac{\sigma^2\lambda_1\lambda_j}{\sigma_\epsilon^2 + \|\boldsymbol{\lambda}_J\|_2^2\sigma^2},$$

it follows that $\tilde{w}_j \geq 0$ if and only if all factor loadings λ_j have the same sign.

For statement 1, Condition (2) implies that given a λ_1 the λ_j lie in a sphere in $\mathbb{R}^J$. A sufficient condition for existence of such λ_j is that the discriminant for the second degree equation for each λ_j is positive. We exemplify this argument for $J = 2$. Condition (2) then requires

$$\lambda_2^2 - \lambda_1\lambda_2 + \sigma_\epsilon^2/(2\sigma^2) + \lambda_3^2 - \lambda_1\lambda_3 + \sigma_\epsilon^2/(2\sigma^2) = 0,$$

which has real roots when the discriminant of each second degree regression is non-negative. In both cases, the condition required is $\lambda_1^2 - 4\sigma_\epsilon^2/(2\sigma^2) \geq 0$. For an arbitrary J it follows that a sufficient condition for real roots is $\lambda_1^2 \geq 4\sigma_\epsilon^2/(J\sigma^2)$.

For statement 2, observe that given that $\|\boldsymbol{\lambda}_J\|_2^2 \geq 0$, we can re-write condition (2) as

$$1 - \lambda_1 \frac{\sum_j \lambda_j}{\|\boldsymbol{\lambda}_J\|_2^2} + \frac{\sigma_\epsilon^2}{\|\boldsymbol{\lambda}_J\|_2^2 \sigma^2} = 0.$$

Given condition (1) it is without loss to write the second term as

$$\frac{\sum_j |\lambda_1 \lambda_j|}{\|\boldsymbol{\lambda}_J\|_2^2} \leq |\lambda_1| \frac{\sum_j |\lambda_j|}{\|\boldsymbol{\lambda}_J\|_2^2} \to 0$$

by the assumption $\frac{1}{\|\boldsymbol{\lambda}_J\|_2^2} \sum_j |\lambda_j| \to 0$. Given that this assumption implies $\|\boldsymbol{\lambda}_J\|_2^2 \to \infty$, as $J \to \infty$ condition (2) is violated as $1 \neq 0$.

For statement 3, recall that an odd function h satisfies that $h(-x) = -h(x)$ which we extend for multidimensional functions to mean this definition simultaneously holds for all components λ_j: $h(-\lambda_2, \ldots, -\lambda_{J+1}) = -h(\lambda_2, \ldots, \lambda_{J+1})$. When h is component-wise weakly increasing, this ensures that for all j, $sign(\lambda_j) = sign(h(\boldsymbol{\lambda}_J)) = sign(\lambda_1)$, so condition (1) is satisfied. As before, if $\|\boldsymbol{\lambda}_J\|_2^2 \to \infty$ condition (2) is satisfied when as $J \to \infty$

$$|\lambda_1| \frac{\sum_j |\lambda_j|}{\|\boldsymbol{\lambda}_J\|_2^2} \to 1,$$

where the λ_1 can be substituted in to get the desired result.

For statement 4, note that we can write the system of equations

$$\boldsymbol{\lambda}_J' \boldsymbol{w} \|\boldsymbol{\lambda}_J\|_1 = \|\boldsymbol{\lambda}_J\|_2^2, \quad \boldsymbol{w} \in \Delta^J,$$

as a linear programming problem. A sufficient condition for this program to have a solution is for $\frac{\|\boldsymbol{\lambda}_J\|_2^2}{\|\boldsymbol{\lambda}_J\|_1}$ to be in the convex hull of the entries of $\boldsymbol{\lambda}_J$.

B.5 Proof of Theorem 4

Recall that under assumptions **A1-A2** by the weak law of large numbers we have that as $T_0 \to \infty$, $\frac{1}{T_0} \sum_t F_t^2 \xrightarrow{p} \sigma^2$, $\frac{1}{T_0} \sum_t F_t \xrightarrow{p} 0$, $\frac{1}{T_0} \sum_t \epsilon_{it}^2 \xrightarrow{p} \sigma_\epsilon^2$ and $\frac{1}{T_0} \sum_t \epsilon_{it} \xrightarrow{p} 0$. For notational convenience drop in what follows the MLE subscript in the weight estimator. By an application of l'hopital rule for the limit, observe that the sum of squares in the log-likelihood l_{T_0} is proportional to

$$
\begin{aligned}
\frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{1t}-\sum_{j=2}^{J+1}\hat{w}_jY_{jt}\right)^2 &= \frac{1}{T_0}\sum_{t=1}^{T_0}(\lambda_1-\sum_{j=2}^{J+1}\hat{w}_j\lambda_j)^2F_t^2 \\
&\quad + 2(\lambda_1-\sum_{j=2}^{J+1}\hat{w}_j\lambda_j)F_t(\epsilon_{1t}-\sum_{j=2}^{J+1}\hat{w}_j\epsilon_{jt}) \\
&\quad + (\epsilon_{1t}-\sum_{j=2}^{J+1}\hat{w}_j\epsilon_{jt})^2 \\
&= \frac{1}{T_0}\sum_{t=1}^{T_0}(\lambda_1-\sum_{j=2}^{J+1}\hat{w}_j\lambda_j)^2\sigma^2+\sigma_\epsilon^2(1+\sum_{j=2}^{J+1}\hat{w}_j^2)+o_p(1).
\end{aligned}
$$

Minimizing the last expression as $T_0 \to \infty$ is equivalent to minimizing the Ridge regression problem, and given that the problem is convex, a global minimizer is given by

$$
\hat{w}_j = \frac{\sigma^2\lambda_1\lambda_j}{\sigma_\epsilon^2+\|\boldsymbol{\lambda}_J\|_2^2\sigma^2}.
$$

Hence, as $T_0 \to \infty$ we have that $\hat{\boldsymbol{w}}_{MLE} \xrightarrow{p} \tilde{\boldsymbol{w}}$. The second statement follows from the conditions given in Theorem 2 and the continuous mapping theorem.

For the third statement, note that under the model assumptions, the standard M-estimator regularity conditions (Vaart (1998)) are satisfied, therefore the result follows from known results. A sketch of the arguments is as follows: by part (1) we have that the estimator is consistent, and by the standard CLT, given that $\mathbb{E}[\nabla_{\boldsymbol{w}} l_{T_0}(\boldsymbol{\theta})|_{\tilde{\boldsymbol{w}}} = 0$, we have that

$$
\sqrt{T_0}\nabla_{\boldsymbol{w}} l_{T_0}(\boldsymbol{\theta})|_{\tilde{\boldsymbol{w}}} \xrightarrow{d} N(0, V),
$$

where $V = V_{T_0}$. The proof then follows from an application of the mean value theorem.

B.6 Proof of Theorem 5

This proof follows results in He and Shao (2000) and He and Shao (1996). A more modern treatment of similar results can also be found in Belloni et al. (2015). The proof considers the more general problem for $M-$estimators in linear models. Start by defining the objective function

$$
\rho(\boldsymbol{w}) = \sum_t (y_{1t}-\boldsymbol{y}_{Jt}'\boldsymbol{w})^2,
$$

and the score function

$$
\phi(\boldsymbol{w}) = 2\sum_t (y_{1t}-\boldsymbol{y}_{Jt}'\boldsymbol{w})\boldsymbol{y}_{Jt}.
$$

Focus on the first assumption:

$$\frac{1}{T_0}\sum_t \boldsymbol{y}_{Jt}\boldsymbol{y}'_{Jt} = D_{T_0},$$

where $0 < \lim\, \inf_{T_0}\sigma_{min}(D_{T_0}) \leq \lim\, \sup_{T_0}\sigma_{max}(D_{T_0}) < \infty$. First note that under A1-A2 the matrix $\boldsymbol{y}_{Jt}\boldsymbol{y}'_{Jt}$ is invertible and has eigenvalues bounded away from zero, so this assumption is satisfied in our setting. Then, under the other assumptions,

$$\left|\boldsymbol{\alpha}'\left(\sum_t \mathbb{E}((y_{1t}-\boldsymbol{y}'_{Jt}\hat{\boldsymbol{w}}_{MLE})-(y_{1t}-\boldsymbol{y}'_{Jt}\tilde{\boldsymbol{w}}))\boldsymbol{y}_{Jt}\right)-T_0\boldsymbol{\alpha}'D_{T_0}(\hat{\boldsymbol{w}}_{MLE}-\tilde{\boldsymbol{w}})\right| \\ \leq c\sum_t|\boldsymbol{y}'_{Jt}\boldsymbol{\alpha}||\boldsymbol{y}'_{Jt}(\hat{\boldsymbol{w}}_{MLE}-\tilde{\boldsymbol{w}})|^2$$

Which implies that there exist a sequence of $J\times J$ matrices D_{T_0} with bounded eigenvalues such that for any $\delta>0$, uniformly in $\boldsymbol{\alpha}\in\mathcal{S}_J(1)$,

$$\sup_{\|w-\tilde{w}\|\leq\delta(J/T_0)^{1/2}}\left|\boldsymbol{\alpha}'\left(\sum_t \mathbb{E}((y_{1t}-\boldsymbol{y}'_{Jt}\boldsymbol{w})-(y_{1t}-\boldsymbol{y}'_{Jt}\tilde{\boldsymbol{w}}))\boldsymbol{y}_{Jt}\right)-T_0\boldsymbol{\alpha}'D_{T_0}(\hat{\boldsymbol{w}}_{MLE}-\tilde{\boldsymbol{w}})\right| \\ = o((T_0J)^{1/2}).$$

The above expression means that condition (C3) in He and Shao (2000) is satisfied. Next, we check that condition (C2) is satisfied. Observe that

$$\|\phi(\tilde{\boldsymbol{w}})\|\leq 2\sum_t\|y_{1t}-\boldsymbol{y}'_{Jt}\tilde{\boldsymbol{w}}\|\|\boldsymbol{y}_{Jt}\| = O((T_0J)^{1/2}),$$

given assumptions (2) and (3) in the theorem. By a similar argument (C1) is satisfied and (C3) implies that (C4) or (C5) are satisfied. Hence, we can apply Corollary 2.1 in He and Shao (2000) to get the desired result.

B.7 Proof of Corollary 6

To derive the first statement in Corollary 5.1 observe that

$$\boldsymbol{y}'_{Jt}(\tilde{\boldsymbol{w}}-\boldsymbol{w})\leq\|\boldsymbol{y}_{Jt}\|^2\|\tilde{\boldsymbol{w}}-\boldsymbol{w}\|^2 = O_p(J^2/T_0),$$

by condition 2 and result 1 in Theorem 5. A similar, approach to He and Shao (2000) can be followed to improve the rate to J/T_0 up to log terms.

To derive the second statement, we show that Theorem 5 applies when α is itself a stochastic sequence. We will show that in our setting the boundedness assumptions on the DGP allows us to preserve stochastic equi-continuity, and so condition (C2) in He and Shao (2000) and in the proof of

Theorem 5 applies. Consider the following quantity

$$\gamma(\boldsymbol{\alpha}) = T_0 \boldsymbol{\alpha}' D_{T_0}(\hat{\boldsymbol{w}}_{MLE} - \tilde{\boldsymbol{w}}).$$

Observe that when $\boldsymbol{\alpha} \in S^J$, a ball of dimension J, by a standard maximal inequality for simplex random variables we get that this object can be bounded by $\sqrt{J T_0}$. It is without loss to substite $\boldsymbol{\alpha}$ with $\boldsymbol{y}_{Jt}$ given assumption 2 in Theorem 5. Indeed, in that case we can also upper bound the quantity by the same rate. Overall, this implies that condition (C2) in He and Shao (2000) is satisfied and so the result follows.

B.8 Proof of Theorem 7

Recall that our Bayes model given an $i.i.d$ sample of size T_0 with $J+1$ units means that the posterior predictive distribution is normal with the following parameters:

1. Mean:
$$\mu^B_{T_0,J} = \frac{\sigma_y^2}{\sigma_y^2 + T_0 \sum_j \tau_j^2} \sum_t \boldsymbol{y}'_{Jt} \mu_J + \frac{\sum_j \tau_j^2}{\sigma_y^2 + T_0 \sum_j \tau_j^2} \sum_t y_{1t}.$$

2. Variance:
$$\Sigma^B_{T_0,J} = \frac{\sigma_y^2 \sum_j \tau_j^2}{\sigma_y^2 + T_0 \sum_j \tau_j^2}.$$

Suppose, as in the assumptions, that $\|\boldsymbol{\mu}_J\|^2 \to 0$ and $\|\lambda_1 - \boldsymbol{\lambda}'_{\boldsymbol{J}} \boldsymbol{\mu}_J\| \to 0$ as $J \to \infty$. Under the DGP given by **A1-A2** we have that $y_{jt} = \lambda_j F_t + \epsilon_{jt}$, for $\epsilon_{jt} \sim_{i.i.d} N(0,1)$. It follows that

$$\begin{aligned}
\boldsymbol{y}'_{JT_0+1} \boldsymbol{\mu}_J &= \sum_j (\lambda_j F_{T_0+1} + \epsilon_{jT_0+1}) \mu_j \\
&= F_{T_0+1} \sum_j \lambda_j \mu_j + \sum_j \epsilon_{jT_0+1} \mu_j.
\end{aligned}$$

Next, consider the following expectation

$$\begin{aligned}
\mathbb{E}[(\boldsymbol{y}'_{JT_0+1} \boldsymbol{\mu}_J - \lambda_1 F_{T_0+1})^2 | F_{T_0+1}] &= \mathbb{E}[(F_{T_0+1}(\sum_j \lambda_j \mu_j - \lambda_1) + \sum_j \epsilon_{jT_0+1} \mu_j)^2 | F_{T_0+1}] \\
&= \mathbb{E}[(F_{T_0+1}^2(\sum_j \lambda_j \mu_j - \lambda_1)^2 + (\sum_j \epsilon_{jT_0+1} \mu_j)^2 | F_{T_0+1}] \\
&\leq \mathbb{E}[(F_{T_0+1}^2(\sum_j \lambda_j \mu_j - \lambda_1)^2 | F_{T_0+1}] + \sigma_\epsilon^2 \sum_j \mu_j^2 \\
&\to 0.
\end{aligned}$$

where the second and third inequalities follow from ϵ_{jt} being mean zero and $i.i.d.$ Under the assumptions on $\boldsymbol{\mu}_J$ it follows that this inequality goes to zero as $J \to \infty$. Therefore, the convergence in probability follows by Chebyshev's inequality.

Next, we show that the mean of the posterior predictive distribution $\mu^B_{T_0,J}$ converges to the same mean as the MLE estimator. First, note that under **A1-A2** and $\sum_j \tau_j^2 = O(T_0^{-\alpha})$ the second term in $\mu^B_{T_0,J}$ is $o_p(1)$ as $\frac{1}{T_0}\sum_t y_{1t} \xrightarrow{p} 0$. The first term then converges to the treated unit factor loading by a similar argument as the first part of this proof given our convex recovery assumption.

To derive the Bernstein-von Mises result we start by noting that due to Pinsker's inequality

$$\|\Phi_{MLE} - Q\|_{TV} \leq \sqrt{\frac{1}{2} D_{KL}(\Phi_{MLE}||Q)}.$$

Hence, we procced in bounding the KL divergence. A useful result is Barron (1986), which provide conditions under which the KL divergence and CLT can be related. The following Lemma summarizes the result.

Lemma 6 (KL Convergence (Barron 1986)). *Let $\Phi_{J,T}$ be the MLE estimator distribution and $Q_{T,J}$ be the smooth, bounded Bayes posterior predictive distribution for fixed J and T_0. Suppose that as $J, T \to \infty$,*

1. *$\Phi_{J,T} \to P^*$,*

2. *$Q_{T,J} \to Q^*$,*

3. *Q^* and P^* have the same mean and have bounded fourth moments.*

Then, it follows that

$$D_{KL}(\Phi_{J,T}||Q_{T,J}) = D_{KL}(\Phi^*||Q^*) + O(1/(TJ)).$$

The conditions in Lemma A.1 are satisfied given our assumptions, the previous derivations and the results in Theorem 5. Therefore, we need to bound $D_{KL}(\Phi^*||Q^*)$. Given our Gaussian assumption we can bound this quantity by comparing the variances of the two distributions. The following Lemma gives an exact derivation of the KL divergence for Gaussian distributions.

Lemma 7 (Gaussian KL). *Suppose that Q and P are normal random variables with equal means and $k \times k$ covariance matrices Σ_Q and Σ_P. Then,*

$$D_{KL}(P||Q) = \frac{1}{2}\left(\log \frac{|\Sigma_P|}{|\Sigma_Q|} - k + tr(\Sigma_Q^{-1}\Sigma_P)\right).$$

Observe that in our case the distributions are one dimensional so it is sufficient to show that as $T_0, J \to \infty$

$$\frac{\mathbb{V}(\Phi^{MLE}_{T_0,J})}{\Sigma^B_{T_0,J}} \to 1.$$

Starting with the Bayesian model, recall that $\sum_j \tau_j^2 = O(J^{\alpha})$, so for the prediction period T_0+1,

$$\Sigma^B_{T_0+1,J} = \frac{\sigma_y^2}{\sigma_y^2(\sum_j \tau_j^2)^{-1}+1} \to \sigma_y^2.$$

From Theorem 5 recall that $\sigma_\alpha^2 = (\mathbb{E}[\epsilon_{Jt}^2])^{-1}\boldsymbol{\alpha}' D_{T_0}^{-1}\boldsymbol{\alpha}$. Following, Corollary 6 when $\boldsymbol{\alpha}$ is given by $\boldsymbol{y}_{Jt}$ and $\frac{1}{T_0}\sum_t \boldsymbol{y}_{Jt}\boldsymbol{y}'_{Jt} = D_{T_0}$ observe that

$$\begin{aligned}
\sigma_\alpha^2 &= (\sigma_\epsilon^2)\boldsymbol{\alpha}' D_{T_0}^{-1}\boldsymbol{\alpha} \\
&= (\sigma_\epsilon^2) T_0 \boldsymbol{y}'_{JT_0+1}(\boldsymbol{y}_{Jt}\boldsymbol{y}'_{Jt})^{-1}\boldsymbol{y}_{JT_0+1} \\
&= (\sigma_\epsilon^2) T_0 \mathrm{tr}((\boldsymbol{y}_{Jt}\boldsymbol{y}'_{Jt})^{-1}\boldsymbol{y}_{JT_0+1}\boldsymbol{y}'_{JT_0+1}) \\
&= (\sigma_\epsilon^2) O(T_0),
\end{aligned}$$

where the last step follows from Theorem 5's assumption 2 regarding the l_2 norm of $\boldsymbol{y}_J t$. This implies that the sample MLE variance is given by $1/T_0\sigma_\alpha^2 \to \sigma_\epsilon^2$. If the time series structure is such that $\mathrm{tr}((\boldsymbol{y}_{Jt}\boldsymbol{y}'_{Jt})^{-1}\boldsymbol{y}_{JT_0+1}\boldsymbol{y}'_{JT_0+1}) = 1$ as $T_0, J \to \infty$, then it follows that a sufficient condition for

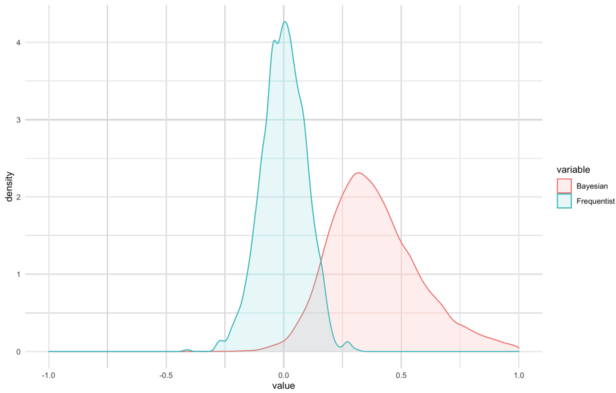
$$D_{KL}(\Phi^*||Q^*) \to 0$$

is that $\sigma_y/\sigma_\epsilon \to 1$ as $J, T \to \infty$. Given **A1 - A2**, it follows that a sufficient condition is $\sigma_y \to \sigma_\epsilon$. The trace condition is satisfied under assumption 2 of Theorem 5 and **A1-A2**, but it would also be satisfied under weaker conditions such as mixing time series regimes.

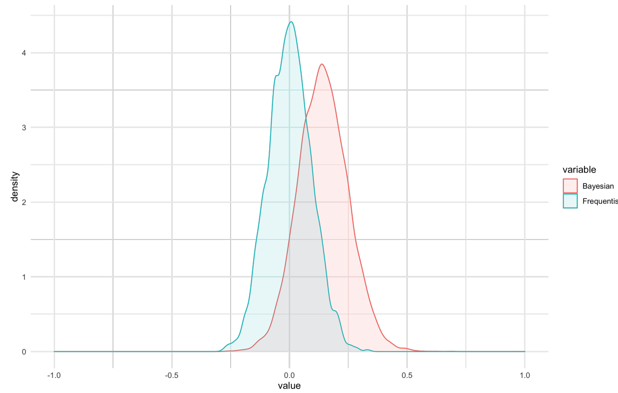
B.9 Additional simulations

In this section we report some additional simulation results in which we compare the frequentist distribution of the treatment effect estimator over 10000 draws with the Bayesian posterior distribution over 1 draw. This simulations are meant to highlight that even with just 1 draw from the underlying DGP as $T_0 \to \infty$ the Bayesian posterior distribution approximates the frequentist distribution. Figure B.2 highlights the BvM convergence in a sparse setting given by the grouped factor model 2.1 in which unit 2 is the only unit to have the same factor loading as unit 1. Figure B.1 replicates Figure B.2 for a denser setting in which we have four groups of five units with equal factor loadings. Unit 1 has the same factor loading as units 2 to 5. The Figure shows that convergence is achieved earlier than in the sparse case, with good coverage around $T = 70$ instead that when $T = 1000$. This is in line with the theory that suggests that a requirement for convergence is that the weights are evenly distributed amongst many units. However, observe that the convergence rate is faster than expected as for 20 units the theory would suggest that at least $20^2 = 400$ time periods are necessary.

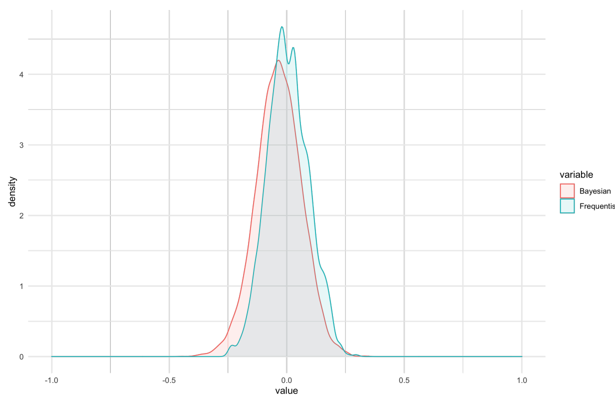
Figure B.1: Convergence of frequentist and Bayesian coverage as $T \to \infty$.



(a) $T = 30$



(b) $T = 50$



(c) $T = 70$

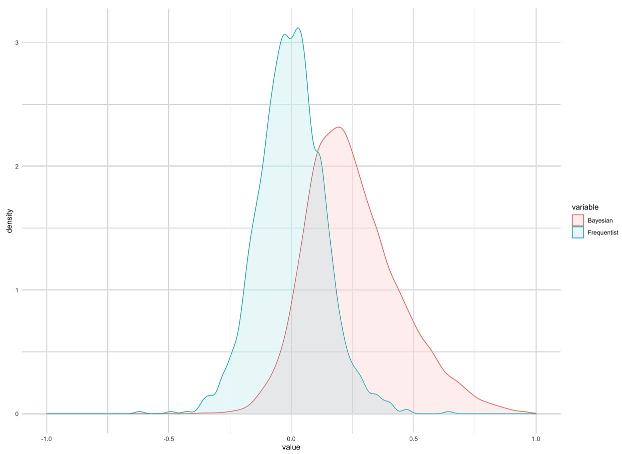
Notes: Kernel densities of the frequentist empirical distribution of the estimated treatment effect over 10000 draws and the Bayesian posterior distribution for one draw for different values of T_0. The potential outcomes are generated by the grouped factor model with 4 groups of 5 units with the same factor loadings and $\sigma = 0.25$.

B.10 German re-unification additional plots

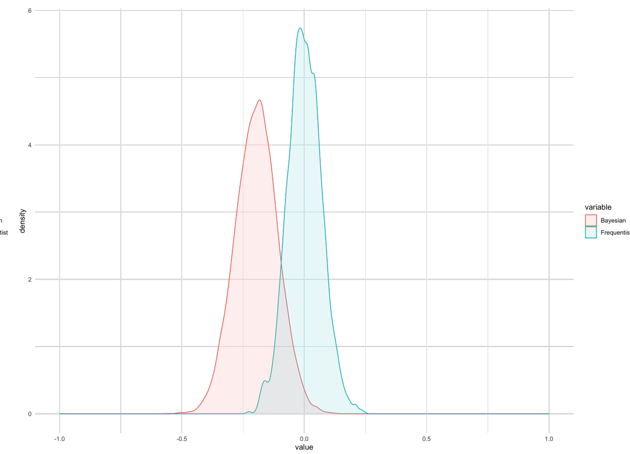
In this section we provide additional materials for the study of the German re-unification. In Figure B.4 we show the correlation between the implicit weight distributions and the distribution of an upper bound on the bias (as proposed by Abadie et al. (2010)) for each Bayesian draw.

B.11 Catalan secession movement additional plots

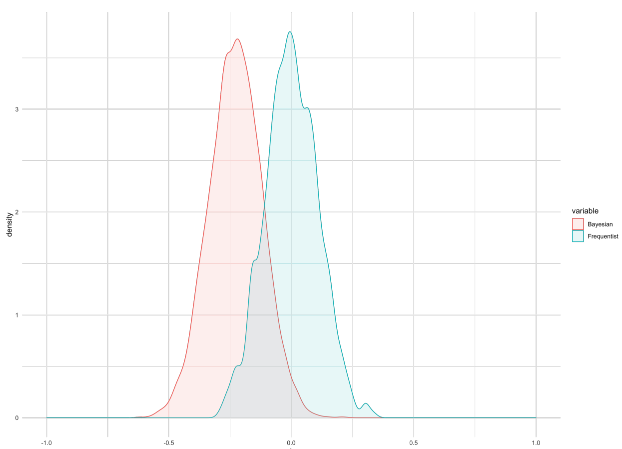
Figure B.2: Convergence of frequentist and Bayesian coverage as $T \to \infty$.



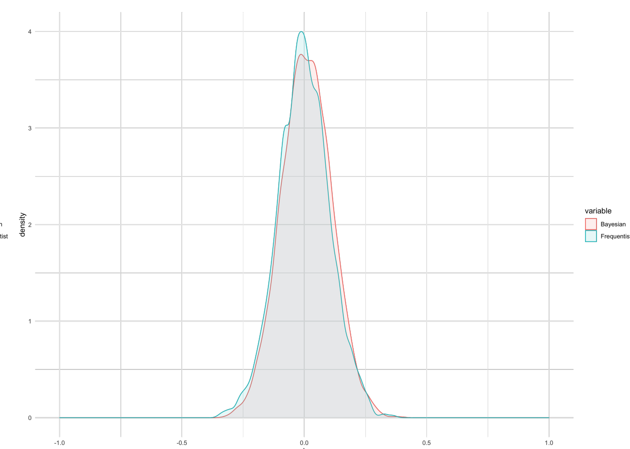
(a) $T = 30$



(b) $T = 100$



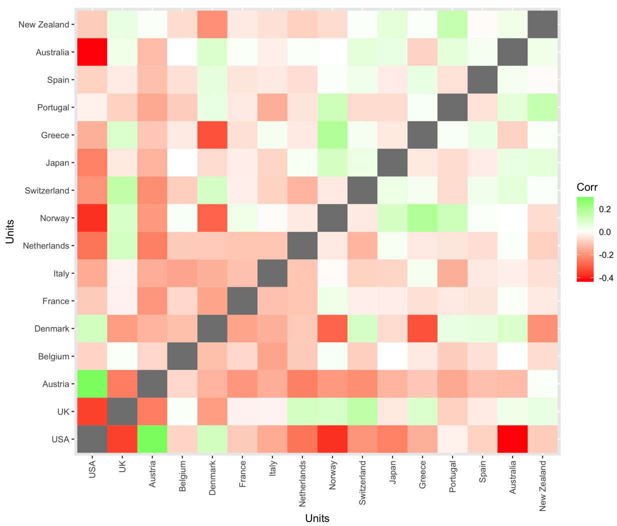
(c) $T = 500$



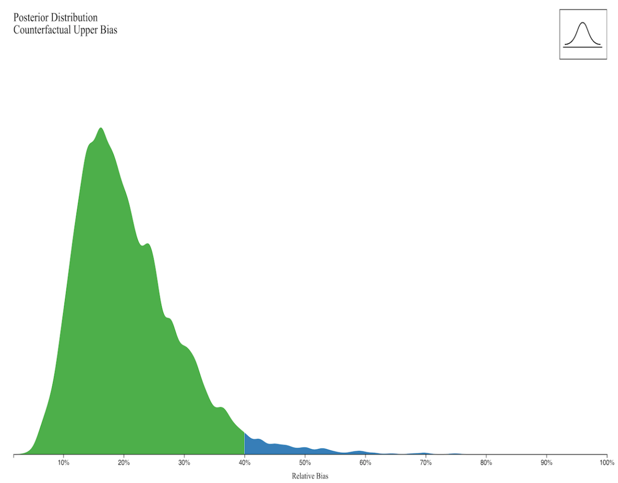
(d) $T = 1000$

Notes: Kernel densities of the frequentist empirical distribution of the estimated treatment effect over 10000 draws and the Bayesian posterior distribution for one draw for different values of T_0. The potential outcomes are generated by the grouped factor model (2.1) with 10 groups of 2 units with $\sigma = 0.25$.

Figure B.3: Additional Plots



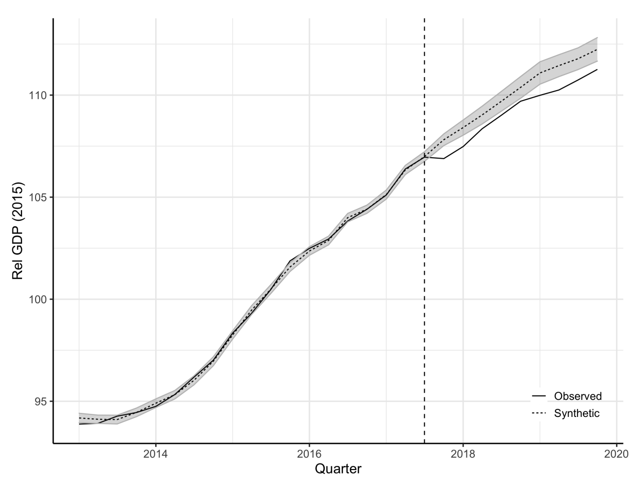
(a) Implicit weights correlations



(b) Bias bound distribution

Notes: Panel (a) shows the correlations between implicit weights of the donor countries. Panel (b) shows the distribution of the bias bound term (computed as the sum of MADs of the pre-treatment outcomes) relative to the size of the mean treatment effect for each Bayesian draw.

Figure B.4: Additional Plot Synthetic Catalonia



Notes: Bayesian synthetic control for Catalonia excluding Madrid and Valencia from the donor pool. The average treatment effect relative to the third quarter of 2017 is -0.85%.

Appendix C

Appendix to Chapter 3

C.1 Note on computational problem

Consider the lower level problem:

$$\psi(\mathbf{V}) = \text{argmin}_{\mathbf{w}\in\mathcal{W}} L_W(\mathbf{V},\mathbf{w}) = \|\mathbf{X}_1^{train} - \mathbf{X}_0^{train}\mathbf{w}\|_V^2.$$

The objective function can be re-written as follows, dropping the *train* label for ease of notation

$$\begin{aligned}\|\mathbf{X}_1 - \mathbf{X}_0\mathbf{w}\|_V^2 &= (\mathbf{X}_1 - \mathbf{X}_0\mathbf{w})'V(\mathbf{X}_1 - \mathbf{X}_0\mathbf{w}) \\ &= \sum_k x_{1k}^2 v_k - 2\sum_j w_j(\sum_k x_{jk}v_k x_{1k}) + \sum_i w_i \sum_k x_{ik}v_k(\sum_j x_{jk}w_j).\end{aligned}$$

Then, the derivative with respect to w_j is given by

$$\frac{\partial L_W}{\partial w_j} = -2(\sum_k x_{jk}v_k x_{1k}) + \sum_k x_{jk}v_k(\sum_i x_{ik}w_i) + \sum_i w_i \sum_k x_{ik}v_k x_{jk}$$

and the derivative with respect to v_k

$$\frac{\partial L_W}{\partial v_k} = x_{1k}^2 - 2\sum_j w_j x_{jk}x_{1k} + \sum_i w_i x_{ik}(\sum_j x_{jk}w_j).$$

Oberve that this implies that while the magnitude of $\psi(\mathbf{V})$ changes with the scale of $\mathbf{V}$ the minimizers $\mathbf{w}$ satisfying $\frac{\partial\psi(\mathbf{V})}{\partial w_j} = 0$ do not. Indeed, multiplying all v_ks by a scalar $a > 0$ will not change the set of minimizers. This implies that we can not jointly minimize the upper and lower level problems directly with the penalty $\lambda\|\text{diag}(\mathbf{V})\|_1$, as $\|\text{diag}(\mathbf{V})\|_1 \to 0$ is optimal when $v_k \geq 0$. This is why in Algorithm 1 we set a $v_k = 1$ before the minimization step.

If $v_{k_0} = 1$ for some $k_0 \in \{1,\ldots,K\}$, then

$$\begin{aligned}\frac{\partial L_W}{\partial w_j} = &-2(x_{jk_0}x_{1k_0} + \sum_{k\neq k_0} x_{jk}v_k x_{1k}) + x_{jk_0}(\sum_i x_{ik}w_i) + \sum_{k\neq k_0} x_{jk}v_k(\sum_i x_{ik}w_i) \\ &+ \sum_i w_i \sum_{k\neq k_0} x_{ik}v_k x_{jk} + \sum_i w_i x_{ik_0}x_{jk_0}.\end{aligned}$$

In this case, the scale of $\mathbf{V}$ is relative to the $\mathbf{X}_{k_0}$, so re-scaling the v_k for $k \neq k_0$ by a positive constant will change the set of minimizers given by $\{\frac{\partial L_W}{\partial w_j} = 0\}$. In practice, the choice of k_0 is meaningful as every k_0 will yield slightly different $\mathbf{w}_j^*$ and the predictor k_0 will be included with probability one in the model. Researchers may choose a k_0 that they know to be important (i.e. a covariate that should be included in the model given domain knowledge), or may treat the k_0 as a hyper-parameter and iterate over all possible ks and choose the k_0 that minimizes the upper level loss.

C.2 Proof of Lemma 1

Consider the linear factor model described in section 3. For notational convenience in this section we do not use boldface terms, and following Abadie et al. (2010) we use Y^P to denote the vector of outcomes for the pre-treatment period. Our counterfactual will be a weighted average of the outcome variable for the donor pool

$$\sum_{j=2}^{J+1} w_j Y_{jt}^N = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \epsilon_{jt}.$$

As a result, the treatment effect $\tau_{1t}^w = Y_{1t} - \sum_{j=2}^{J+1} w_j Y_{jt}^N$ for some weight vector $w = \{w_j\}_{j=2}^{J+1}$ takes the form

$$\tau_{1t}^w = \theta_t \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right) + \lambda_t \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) + \sum_{j=2}^{J+1} w_j (\epsilon_{1t} - \epsilon_{jt}).$$

Under some additional definitions and mean independence of the error term we have that:

$$\begin{aligned}
\tau_{1t}^w = {} & \lambda_t (\lambda^{P\prime} \lambda^P)^{-1} \lambda^{P\prime} \left(Y_1^P - \sum_{j=2}^{J+1} w_j Y_j^P \right) \\
& + \left(\theta_t - \lambda_t (\lambda^{P\prime} \lambda^P)^{-1} \lambda^{P\prime} \theta^P \right) \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right) \\
& - \lambda_t (\lambda^{P\prime} \lambda^P)^{-1} \lambda^{P\prime} \left(\epsilon_1^P - \sum_{j=2}^{J+1} w_j \epsilon_j^P \right) \\
& + \sum_{j=2}^{J+1} w_j (\epsilon_{1t} - \epsilon_{jt}).
\end{aligned}$$

In Abadie et al. (2010), a bias bound is derived in the case in which we perfectly replicate the treated unit and the first two terms are zero. They show that the last two terms are $\mathrm{O}\left(\frac{1}{T_0}\right)$. This paper explores the case in which we have many covariates and therefore are likely to fall outside the convex hull of the donor pool. In such cases, the synthetic control will not be able to replicate the design matrix of the treated unit and the pre-treatment fit will not be perfect.

Assume that $\sum_{t=1}^{T_0} \lambda_t' \lambda_t$ is positive semi-definite with smallest eigenvalue bounded away from zero by $\underline{\xi}$ and $|\lambda_{tf}| < \bar{\lambda}$, $|\theta_{tf}| < \bar{\theta}$ for all t, f. Then, it can be shown by the C-S inequality that:

$$\left(\lambda_t \left(\sum_{s=1}^{T_0} \lambda_t' \lambda_t \right)^{-1} \lambda_m' \right)^2 \leq \left(\frac{\bar{\lambda}^2 F}{T_0 \underline{\xi}} \right)^2.$$

Then, consider the first term in the decomposition of the treatment effect,

$$
\begin{aligned}
\lambda_t(\lambda^{P'}\lambda^P)^{-1}\lambda^{P'}\left(Y_1^P-\sum_{j=2}^{J+1}w_jY_j^P\right) &= \sum_{m=1}^{T_0}\lambda_t\left(\sum_{s=1}^{T_0}\lambda_t'\lambda_t\right)^{-1}\lambda_m'(Y_{1m}-\sum_{j=2}^{J+1}w_jY_{jm})\\
&\leq \sum_{m=1}^{T_0}\left|\lambda_t\left(\sum_{s=1}^{T_0}\lambda_t'\lambda_t\right)^{-1}\lambda_m'\right|\left|Y_{1m}-\sum_{j=2}^{J+1}w_jY_{jm}\right|\\
&\leq \left(\frac{\bar{\lambda}^2F}{T_0\underline{\xi}}\right)\sum_{m=1}^{T_0}\left|Y_{1m}-\sum_{j=2}^{J+1}w_jY_{jm}\right|\\
&= \left(\frac{\bar{\lambda}^2F}{\underline{\xi}}\right)\mathrm{MAD}\left(Y_1^P,\sum_{j=2}^{J+1}w_jY_j^P\right).
\end{aligned}
$$

Therefore, the bias contribution from the first term in the τ_{1t}^w decomposition (which we denote by R_{1t}) is bounded by:

$$
\mathbb{E}|R_{1t}| \leq \left(\frac{\bar{\lambda}^2F}{\underline{\xi}}\right)\mathbb{E}\mathrm{MAD}\left(Y_1^P,\sum_{j=2}^{J+1}w_jY_j^P\right).
$$

Similar, the second term is given by

$$
\begin{aligned}
\left(\theta_t-\lambda_t(\lambda^{P'}\lambda^P)^{-1}\lambda^{P'}\theta^P\right)\left(Z_1-\sum_{j=2}^{J+1}w_jZ_j\right) &= \sum_{k=1}^{K}\left(\theta_{tk}-\sum_{l=1}^{T_0}\lambda_t\left(\sum_{s=1}^{T_0}\lambda_s'\lambda_s\right)^{-1}\lambda_l'\theta_{lk}\right)\left(Z_{1k}-\sum_{j=2}^{J+1}w_jZ_{jk}\right)\\
&\leq \sum_{k=1}^{K}\left|\theta_{tk}-\sum_{l=1}^{T_0}\lambda_t\left(\sum_{s=1}^{T_0}\lambda_s'\lambda_s\right)^{-1}\lambda_l'\theta_{lk}\right|\left|Z_{1k}-\sum_{j=2}^{J+1}w_jZ_{jk}\right|\\
&\leq \left|\bar{\theta}\left(1-\left(\frac{\bar{\lambda}^2F}{T_0\underline{\xi}}\right)\right)\right|\sum_{k=1}^{K}\left|Z_{1k}-\sum_{j=2}^{J+1}w_jZ_{jk}\right|\\
&= K\left|\bar{\theta}\left(1-\left(\frac{\bar{\lambda}^2F}{T_0\underline{\xi}}\right)\right)\right|\mathrm{MAD}\left(Z_1,\sum_{j=2}^{J+1}w_jZ_j\right).
\end{aligned}
$$

The bias contribution from the second term (R_{2t}) is then given by:

$$
\mathbb{E}|R_{2t}| \leq K\left|\bar{\theta}\left(1-\left(\frac{\bar{\lambda}^2F}{T_0\underline{\xi}}\right)\right)\right|\mathbb{E}\mathrm{MAD}\left(Z_1,\sum_{j=2}^{J+1}w_jZ_j\right).
$$

Given that the other two terms are $\mathrm{O}\left(\frac{1}{T_0}\right)$ the result follows.

C.2.1 Proof of Theorem 1

Start by re-writing the bi-level optimization problem as a convex optimization problem where the lower level problem solutions are characterized by the necessary Karush-Kuhn-Tucker (KKT) conditions. The program can therefore be re-written as

$$\begin{aligned}
\min_{\boldsymbol{w},\boldsymbol{v},\kappa_1^w,\boldsymbol{\kappa}_2^w} \quad & \frac{1}{T_{val}}\|\mathbf{Y}_1^{val} - \mathbf{Y}_0^{val}\boldsymbol{w}\|^2 + \lambda(1 + \sum_{k\neq k_0}^{K} v_k) \\
s.t. \quad & 2\mathbf{X}_0^{train'}\mathbf{V}\mathbf{X}_0^{train}\mathbf{w} - 2\mathbf{X}_0^{train'}\mathbf{V}\mathbf{X}_1^{train} + \kappa_1^w\mathbf{1}_J + \boldsymbol{\kappa}_2^w = \mathbf{0}, \\
& \sum_j \kappa_{2j}^w w_j = 0, \\
& \kappa_{2j}^w \geq 0 \text{ for all } j, \\
& v_k \geq 0 \text{ for all } k \neq k_0
\end{aligned}$$

where κ_1^w is the Lagrangian multiplier for the sum to 1 constraint and $\boldsymbol{\kappa}_2^w$ are the associated multipliers for the non-negative constraint for the w_j terms in the inner problem.
The associated Lagrangian for parameter vector $\boldsymbol{\nu} = (\boldsymbol{w}, \boldsymbol{v}, \kappa_1^w, \boldsymbol{\kappa}_2^w, \boldsymbol{\kappa}^v, \boldsymbol{\xi}_1, \xi_2, \boldsymbol{\xi}_3)'$ is

$$\begin{aligned}
\min_{\boldsymbol{\nu}} \mathcal{L}(\boldsymbol{\nu}) \equiv \quad & \frac{1}{T_{val}}\|\mathbf{Y}_1^{val} - \mathbf{Y}_0^{val}\boldsymbol{w}\|^2 + \lambda(1 + \sum_{k\neq k_0}^{K} v_k) \\
& + \boldsymbol{\xi}_1'(2\mathbf{X}_0^{train'}\mathbf{V}\mathbf{X}_0^{train}\mathbf{w} - 2\mathbf{X}_0^{train'}\mathbf{V}\mathbf{X}_1^{train} + \kappa_1^w\mathbf{1}_J + \boldsymbol{\kappa}_2^w) \\
& + \xi_2 \sum_j \kappa_{2j}^w w_j \\
& + \xi_3 \boldsymbol{\kappa}_2^w \\
& + \boldsymbol{\kappa}^{v'}\boldsymbol{v}.
\end{aligned}$$

subject to the KKT dual feasibility and slackness conditions:

$$\begin{aligned}
\textbf{Complementary slackness}: \quad & \sum_{k\neq k_0} \kappa_k^v v_k = 0, \\
& \sum_j \xi_{3j}\kappa_j^w = 0. \\
\textbf{Dual feasibility}: \quad & \boldsymbol{\kappa}^v \geq \mathbf{0}, \quad \boldsymbol{\xi}_3 \geq \mathbf{0},
\end{aligned}$$

where κ_1^v $\boldsymbol{\kappa}_2^v$ are the associated Lagrangian multipliers for the constraints on v_k and $\boldsymbol{\xi}_1, \xi_2$, and $\boldsymbol{\xi}_3$ are the multipliers associated with the lower level KKT necessary conditions.

Observe that the lower level loss first order condition is given by:

$$\mathbf{X}_0^{train'}\mathbf{V}\mathbf{X}_0^{train}\mathbf{w} - \mathbf{X}_0^{train'}\mathbf{V}\mathbf{X}_1^{train} = \sum_k \mathbf{x}_k\mathbf{x}_k' v_k \cdot \mathbf{w} - \sum_k \mathbf{x}_k x_{1k} v_k.$$

Taking the derivative with respect to v_k yields

$$\frac{\partial}{\partial v_k} \implies \mathbf{x}_k(\mathbf{x}_k'\mathbf{w} - x_{1k}).$$

It follows that the derivative of the objective function of $\mathcal{L}(\boldsymbol{\nu})$ with respect to v_k is given by

$$\frac{\partial \mathcal{L}(\boldsymbol{\nu})}{\partial v_k} = \xi_1'\mathbf{x}_k(\mathbf{x}_k'\mathbf{w} - x_{1k}) + \kappa_k^v + \lambda.$$

Under the assumption that that $\psi : \mathcal{V} \to \mathcal{W}$ is an injective function, it follows that there is a unique solution to the lower level optimization problem. This is the case when every subset of k columns of $\boldsymbol{X}_0$ spans $\mathbb{R}^k$ (that is, the columns in every subset are linearly independent) and when $\boldsymbol{V}$ is diagonal and positive definite so that $\boldsymbol{X}_0'\boldsymbol{V}\boldsymbol{X}_0$ is non-singular. Since $\lambda, \kappa_k^v \geq 0$, the FOC implies that $\rho_k \equiv \xi_1'\mathbf{x}_k(\mathbf{x}_k'\mathbf{w} - x_{1k}) \leq 0$. Given a choice of λ, by the complementary slackness condition $\kappa_v^k > 0$ for all k such that $\rho_k + \lambda = 0$ and $\kappa_v^k = 0$ otherwise. Therefore, the choice of λ, together with $\boldsymbol{w}$ and the multipliers ξ_1, control how many predictors will receive zero weight.

We will now show that as $T_0, T_{val} \to \infty$, the ρ_k for $k \in S = \{k \mid \theta_{kt} = 0 \text{ for all } k, t\}$, do not converge to zero in probability. Consider using the predictors Z for the design matrix $\boldsymbol{X}$, such that $\rho_k = \xi_1'\mathbf{x}_k(\mathbf{x}_k'\mathbf{w} - x_{1k}) = \xi_1'\mathbf{Z}_k(\mathbf{Z}_k'\mathbf{w} - Z_{1k})$. Recall the separation assumption,

$$\boldsymbol{w}^* \in \operatorname{argmin}_{w \in \Delta^J} \mathbb{E}\|\mathbf{Y}_1^{val} - \mathbf{Y}_0^{val}\boldsymbol{w}\|^2$$

satisfies that for all $k \in S$ and $l \in S^c$, $|Z_{1k} - Z_{Jk}'\boldsymbol{w}^*| > 0$. By the reverse triangle inequality

$$\begin{aligned} |Z_{1k} - Z_{Jk}'\boldsymbol{w}| &= |Z_{1k} - Z_{Jk}'\boldsymbol{w}^* - Z_{Jk}'(\boldsymbol{w} - \boldsymbol{w}^*)| \\ &\geq \big||Z_{1k} - Z_{Jk}'\boldsymbol{w}^*| - |Z_{Jk}'(\boldsymbol{w} - \boldsymbol{w}^*)|\big| \\ &\geq |Z_{1k} - Z_{Jk}'\boldsymbol{w}^*| - \sqrt{k}\sum_j (w_j - w_j^*). \end{aligned}$$

given that by assumption we have that $\|Z_j\|^2 \leq \sqrt{k}$. It follows that if for fixed J $\boldsymbol{w} \xrightarrow{p} \boldsymbol{w}^*$ as $T_{val} \to \infty$ then $P(|Z_{1k} - Z_{Jk}'\boldsymbol{w}| > 0) \to 1$ as $T_{val} \to \infty$ so we don't have convergence in probability. Furthermore, under the injective assumption of ψ the Z are linearly independent so it follows that $\xi_1'\mathbf{Z}_k \neq 0$ as by feasibility and the weight sum to one restriction ξ_{1j} can not be zero for all j. Therefore, it follows that for fixed J, if as $T_0, T_{val} \to \infty$, $\boldsymbol{w} \xrightarrow{p} \boldsymbol{w}^*$ and $\lambda \to 0$, then

$$P(\rho_k = 0) \to 0.$$

This implies that $P(\kappa_k^v = 0) \to 0$ and, therefore, that for $k \in S$ for fixed J, if as $T_0, T_{val} \to \infty$, $\boldsymbol{w} \xrightarrow{p} \boldsymbol{w}^*$ and $\lambda \to 0$, then

$$P(v_k = 0) \to 1.$$

This shows the first part of the theorem. If, furthermore, we have that for $l \in S^c$ $|Z_{1k} - Z'_{Jk}\boldsymbol{w}^*| = 0$ (i.e. perfect fit) then a similar argument yields that for fixed J, if as $T_0, T_{val} \to \infty$, $\boldsymbol{w} \xrightarrow{p} \boldsymbol{w}^*$ and $\lambda \to 0$, then

$$P(v_l = 0) \to 0.$$

Next, we show that for fixed J we have that $\boldsymbol{w} \xrightarrow{p} \boldsymbol{w}^*$ as $T_0, T_val \to \infty$. Let $Q = \mathbb{E}\|\boldsymbol{Y}_1^{val} - \boldsymbol{Y}_0^{val}\boldsymbol{w}\|^2$ and $\hat{Q} = \frac{1}{T_{val}}\|\boldsymbol{Y}_1^{val} - \boldsymbol{Y}_0^{val}\boldsymbol{w}\|^2$. Given the covariance stationarity assumption by a suitable LLN it follows that for any w we have that $\tilde{Q} \xrightarrow{p} Q$ as $T_{val} \to \infty$. Hence, it follows that $\sup_w |\tilde{Q}(\boldsymbol{w}) - Q(\boldsymbol{w})| \to 0$ as $T_{val} \to \infty$. Furthermore, we have that $\tilde{Q}$ and Q are continuous and convex, and the simplex Δ^J is compact. Therefore, by Lemma 3 of Amemiya (1973) (or Theorem 2.1 in Newey and McFadden (1994)) it follows that $\tilde{w} \xrightarrow{p} \boldsymbol{w}^*$ where $\tilde{w} \in \text{argmin}_{w \in \Delta^J} \hat{Q}(\boldsymbol{w})$.

In general $\tilde{w}$ need not be equivalent to the $\boldsymbol{w}$ that solves the bi-level program. However, as $T_0, T_{val} \to \infty$ and λ is chosen to minimize the validation loss as in our bilevel program, $\tilde{w} \xrightarrow{p} \tilde{w}$. The main observation is that the bilevel loss nests the $\tilde{Q}$ loss at the optimum. Indeed, set all v_k weights to 0 and $\lambda = 0$, then the lower level problem loss depends on $|Z_{1k_0} - Z'_{Jk_0}w|$ and the upper level problem loss is equivalent to $\tilde{Q}$. Under the oracle separation assumption it follows that we only need to consider the upper level loss as it will simultaneously minimize the lower level problem. For instance, if we can perfectly fit the predictors $k \in S^c$ it follows that as $T_0, T_{val} \to \infty$ our bilevel loss coincides with $\tilde{Q}$ therefore it follows by the trivial inequality (can't do better than $\boldsymbol{w}^*$) that $\boldsymbol{w} \xrightarrow{p} \boldsymbol{w}^*$.

C.3 Proof of Theorem 2: MSE Rates

We focus on the covariate matching problem with subGaussian noise:

$$Z_1 = Z_0 w^* + \epsilon, \quad \epsilon \sim_{ind} \text{subG}(\sigma_z^2).$$

Under A1-A3 we assume a sparse representation where only k_1 predictors are non-zero. Theorem 1 implies that as $T_0 \to \infty$ almost surely the sparse predictors are not used in the model. Hence, with probability one, we can re-write the optimization problem in terms of the true underlying sparse model:

$$
\begin{aligned}
min_{V,W} L_V(V,W) &= \frac{1}{T_{val}} \|Y_1^{val} - Y_0^{val} W(V)\|^2 + \lambda \|V\|_1, \\
\text{s.t. } & W(V) \in \psi(V), \\
& V \in \mathcal{V},
\end{aligned}
$$

where $\psi : \mathcal{V} \rightrightarrows \mathcal{W}$ maps the upper level solutions to the lower level optima

$$
\psi(V) \equiv argmin_{W \in \mathcal{W}} L_W(V,W) = \|X_1^{train} - X_0^{train} W\|_V^2,
$$

and

$$
\begin{aligned}
W \in \mathcal{W} &\equiv \left\{ W \in \mathbb{R}^J \mid \mathbf{1}'W = 1,\ W_j \geq 0,\ j = 2, \ldots, J+1 \right\} \equiv \Delta^J, \\
V \in \mathcal{V} &\equiv \{(v_1, \ldots, v_k) \mid v_j = 0 \text{ for } j \in S,\ v_j \in \mathbb{R} \text{ for } j = 1, \ldots k\}.
\end{aligned}
$$

where $S = \{k \mid \theta_{tk} = 0 \text{ for all } t\}$ denotes the set of nuissance predictors, and $|S^c| = k_1$. Consider the lower level program of matching the covariates. For simplicity, restrict design matrix to the covariates, without including transformations of the outcome variable. Furthermore, let Z be bounded such that $\max_j \|Z_j\| \leq \sqrt{k}$ as in A1-A3. Denote the minimizer of the lower level program by $\hat{w}$, then it follows that

$$
\begin{aligned}
\|Z_1 - Z_0 \hat{w}\|_V^2 &\leq \|Z_1 - Z_0 w^*\|_V^2 \\
\|Z_0 w^* - Z_0 \hat{w}\|_V^2 &\leq 4 \langle V^T \epsilon, \frac{Z_0 w^* - Z_0 \hat{w}}{\|Z_0 w^* - Z_0 \hat{w}\|_V} \rangle^2 \\
&\leq \sup_{b \in Z_0 \Delta^J, \|b\| = 1} 4 \langle V^T \epsilon, b \rangle^2
\end{aligned}
$$

Given our assumptions it can be shown that $Z_j^T V^T \epsilon \sim \text{subG}(k_1 \sigma_z^2)$. Therefore, using a maximal inequality it follows that

$$
MSE(Z_0 \hat{w}) = \frac{1}{k} \mathbb{E} \max_b \langle V^T \epsilon, b \rangle \lesssim \frac{\sigma_z \sqrt{k_1}}{k} \sqrt{2 \log J}.
$$

The rate for the standard synthetic control can be derived in a similar way, without restricting the model to be in the non-sparse support of the predictors.

C.4 Placebo variance estimation

We estimate the noise level using the placebo variance estimation bootstrap proposed in Arkhangelsky et al. (2021), which follows from the placebo exercises first described by Abadie et al. (2010). This bootstrap procedure is valid when the noise distribution of the treated and donor units is the same

(homoskedasticity). A further discussion on similar bootstrap procedures when the number of treated units is small can be found in Conley and Taber (2011).
The bootstrap procedure works as follows. For B replications:

1. Randomly choose a donor unit l to be treated.
2. Compute the synthetic control weights $\hat{\boldsymbol{w}}^b$ using the remaining controls units as the donor pool.
3. Estimate the average treatment effect on the treated in the post treatment period

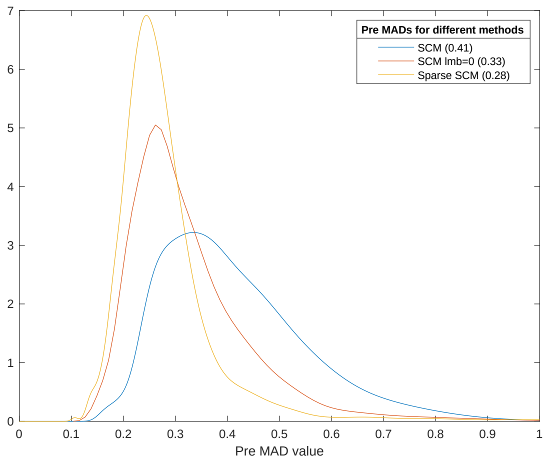
$$\hat{\tau}_b = \frac{1}{T - T_0} \sum_t Y_{lt} - \boldsymbol{Y}'_{-lt} \hat{\boldsymbol{w}}^b.$$

Then the placebo variance estimator is given by

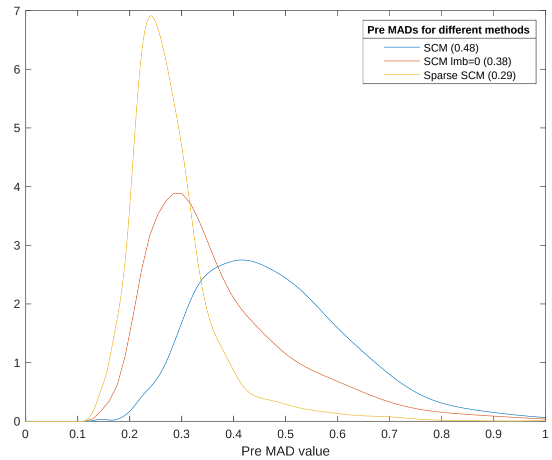
$$\hat{V}_\tau = \frac{1}{B} \sum_b (\hat{\tau}_b - \bar{\tau}_b)^2,$$

where $\bar{\tau}_b = \frac{1}{B} \sum_b \hat{\tau}_b$. In a simulation exercise, Arkhangelsky et al. (2021) show that the coverage properties of the placebo variance bootstrap procedure are good.

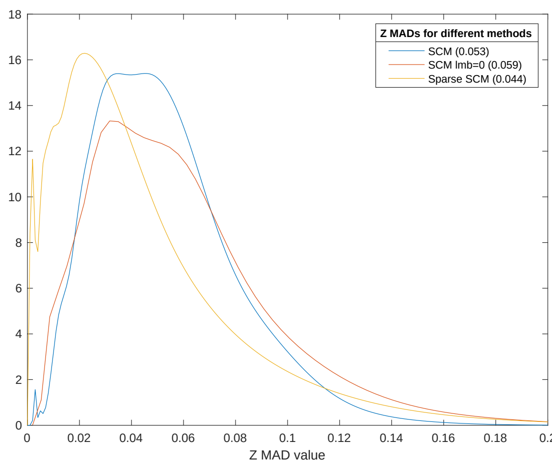
C.5 Additional Simulation Figures



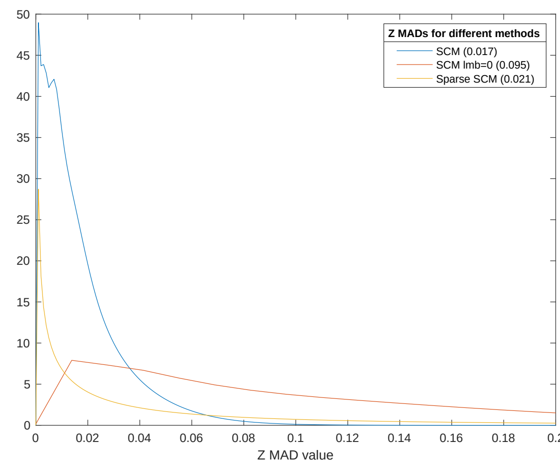
(c) $k_1 = k_2 = 5$.



(d) $k_1 = 1$, $k_2 = 9$.



(e) SCM $\lambda^* = 0$ $\boldsymbol{V}^*$



(f) *Sparse* SCM $\boldsymbol{V}^*$

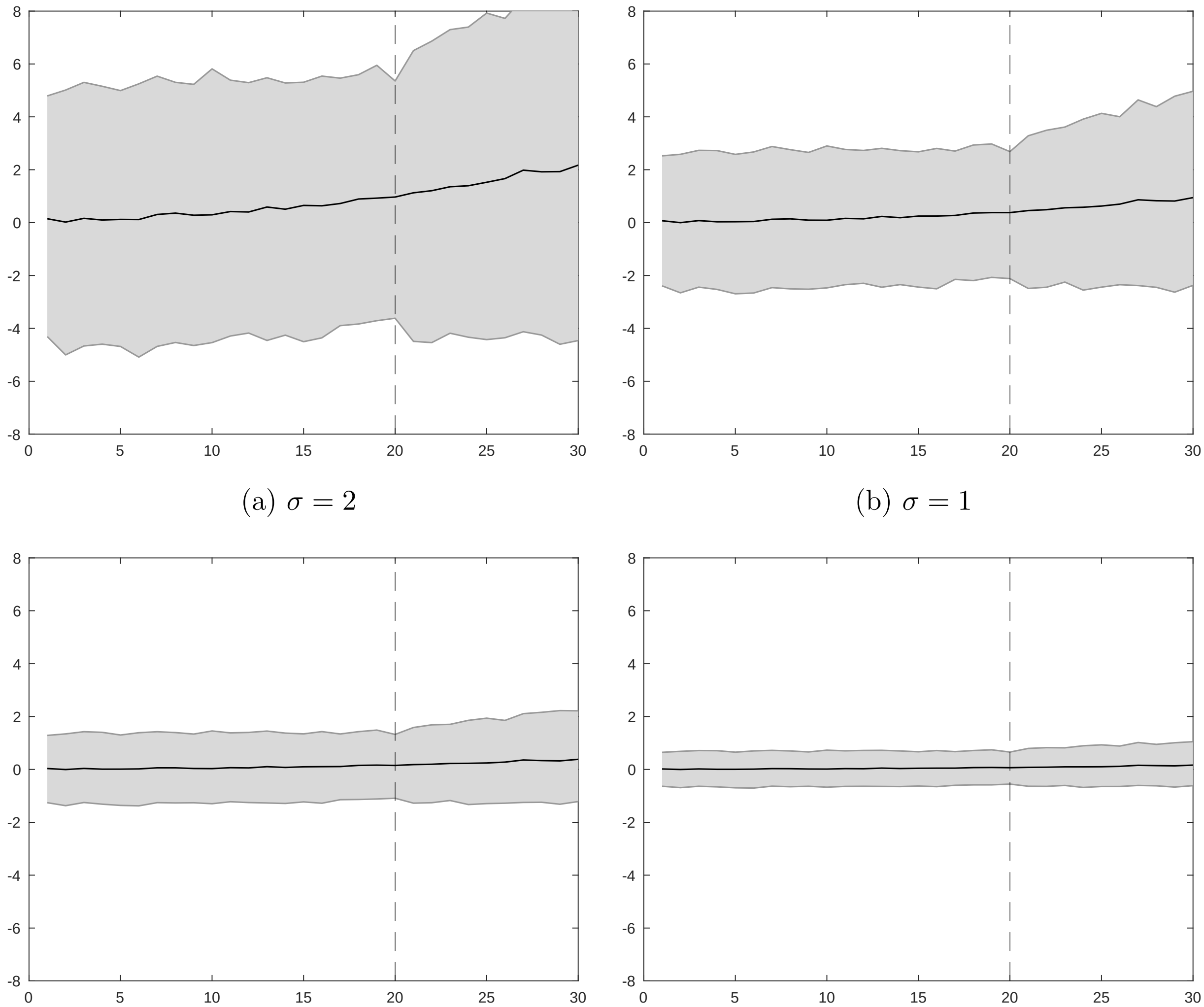
Notes: Panels (a) - (b) show the kernel density across simulations of MADs of the outcome variable for the pre-treatment period, with average values in parenthesis. Panels (c) - (d) show the kernel density across simulations of MADs of the useful predictors, with average values in parenthesis.

Appendix D

Appendix to Chapter 4

Table A.1 summarizes the results for all figures by providing the average pre-RMSE, post-RMSE and W_2 (the weight assigned to the second unit) over 10000 simulations. Figures A.4-4.12 show 95% bands over 10000 simulations of the estimation error (prediction minus actual) for the same simulation designs as in Figures 4.4-4.12, respectively.

Figure A.4: Synthetic control error simulations with a stochastic trend.



(a) $\sigma = 2$ (b) $\sigma = 1$

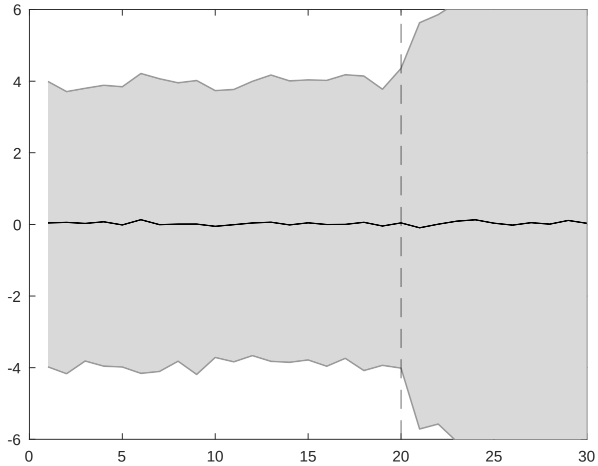
(c) $\sigma = 0.5$ (d) $\sigma = 0.25$

Note: 95% bands for the simulation design of Figure 4.4.

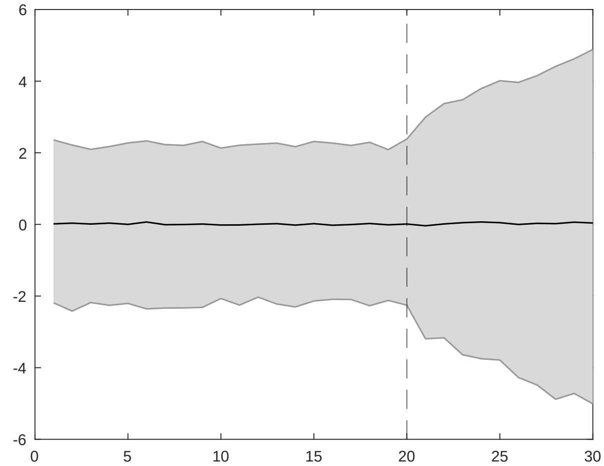
Table A.1: Simulation results for all figures.

Figure	J	T_0	σ	ρ	post-RMSE	pre-RMSE	W_2^*
Figure 4.2 (a)	19	20	0.25	0.5	0.365	0.315	0.890
Figure 4.2 (b)	19	20	0.5	0.5	0.723	0.588	0.730
Figure 4.2 (c)	19	20	1	0.5	1.362	1.032	0.429
Figure 4.2 (d)	19	20	2	0.5	2.421	1.782	0.171
Figure 4.4 (a)	19	20	0.25	1	0.387	0.341	0.983
Figure 4.4 (b)	19	20	0.5	1	0.795	0.675	0.953
Figure 4.4 (c)	19	20	1	1	1.642	1.315	0.882
Figure 4.4 (d)	19	20	2	1	3.272	2.513	0.763
Figure 4.5 (a)	19	20	0.25	1	0.439	0.320	0.920
Figure 4.5 (b)	19	20	0.5	1	0.921	0.609	0.799
Figure 4.5 (c)	19	20	1	1	1.799	1.108	0.564
Figure 4.5 (d)	19	20	2	1	3.117	1.965	0.316
Figure 4.6 (a)	19	4	0.5	1	2.595	0.282	0.397
Figure 4.6 (b)	199	4	0.5	1	3.305	0.057	0.080
Figure 4.6 (c)	19	20	0.5	1	0.921	0.609	0.799
Figure 4.6 (d)	199	20	0.5	1	1.308	0.484	0.529
Figure 4.7 (a)	19	15	0.25	0.5	0.381	0.304	0.874
Figure 4.7 (b)	19	15	0.25	0.5	0.405	0.289	0.860
Figure 4.7 (c)	19	15	0.25	0.5	1.019	0.000	
Figure 4.8 (a)	19	20	0.25	0.5	0.415	0.356	0.841
Figure 4.8 (b)	19	20	0.5	0.5	0.829	0.692	0.631
Figure 4.8 (c)	19	20	1	0.5	1.515	1.234	0.320
Figure 4.8 (d)	19	20	2	0.5	2.579	2.114	0.131
Figure 4.9 (a)	19	20	0.25	0.5	1.694	0.356	0.841
Figure 4.9 (b)	19	20	0.25	0.5	1.682	0.315	0.890
Figure 4.10 (d)	199	20	0.25	1	0.495	0.315	0.875
Figure 4.10 (c)	199	20	0.25	1	0.524	0.305	0.853
Figure 4.10 (b)	199	20	0.25	1	0.562	0.291	0.826
Figure 4.10 (a)	199	20	0.25	1	0.637	0.274	0.786
Figure 4.11 (a)	1000	4	5	1	38.153	0.661	
Figure 4.11 (b)	1000	4	5	1	34.967	1.080	
Figure 4.11 (c)	1000	4	5	1	26.531	1.250	
Figure 4.11 (d)	1000	4	5	1	20.531	1.438	
Figure 4.11 (e)	1000	4	5	1	12.776	1.675	
Figure 4.11 (f)	1000	4	5	1	7.616	1.811	
Figure 4.12	49	20	1		4.093	2.073	

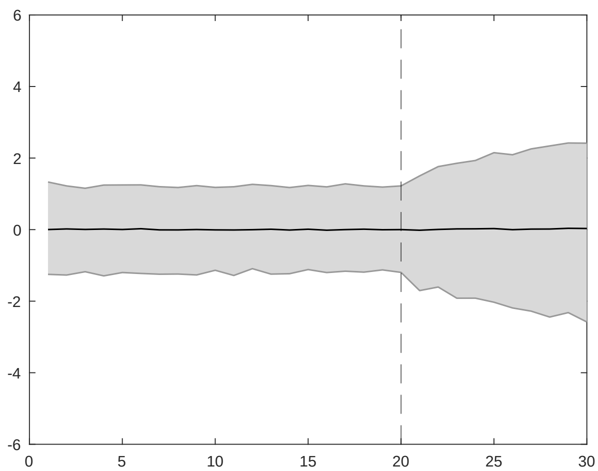
Figure A.5: Synthetic control error simulations when $\rho = 1$.



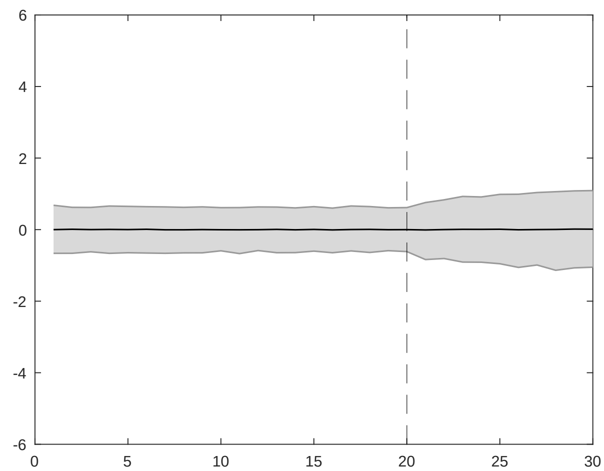
(a) $\sigma = 2$



(b) $\sigma = 1$



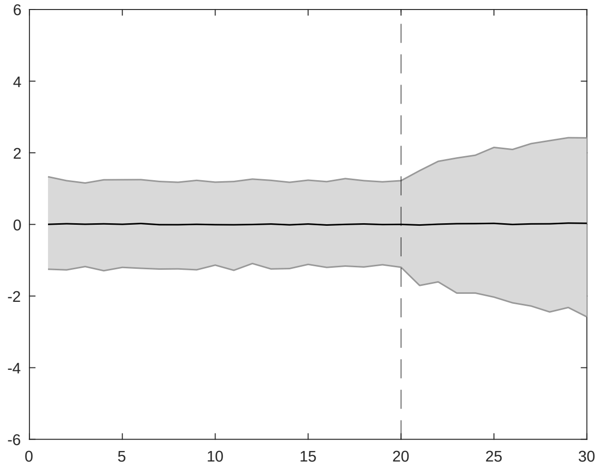
(c) $\sigma = 0.5$



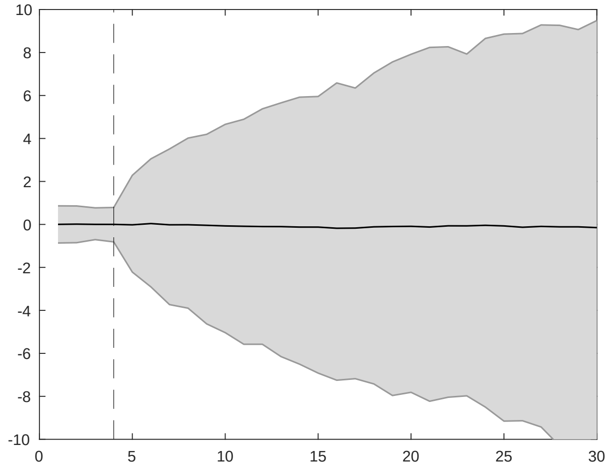
(d) $\sigma = 0.25$

Note: 95% bands for the simulation design of Figure 4.5.

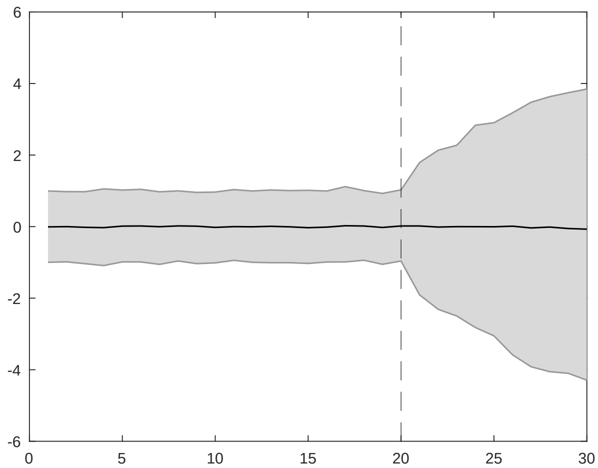
Figure A.6: Synthetic control error simulations and over-fitting.



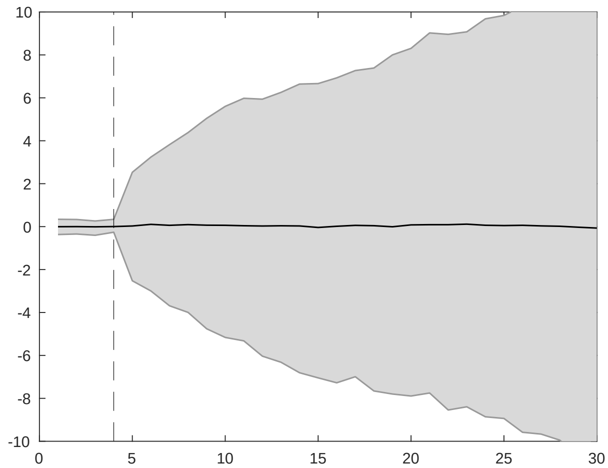
(a) $T_0 = 20$, $F = 10$ ($J = 19$)



(b) $T_0 = 4$, $F = 10$ ($J = 19$)



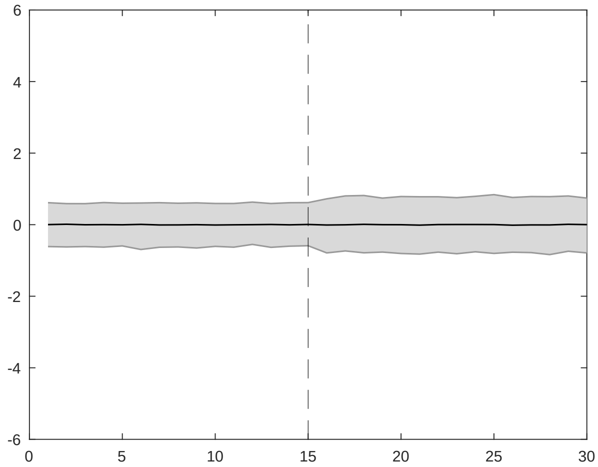
(c) $T_0 = 20$, $F = 100$ ($J = 199$)



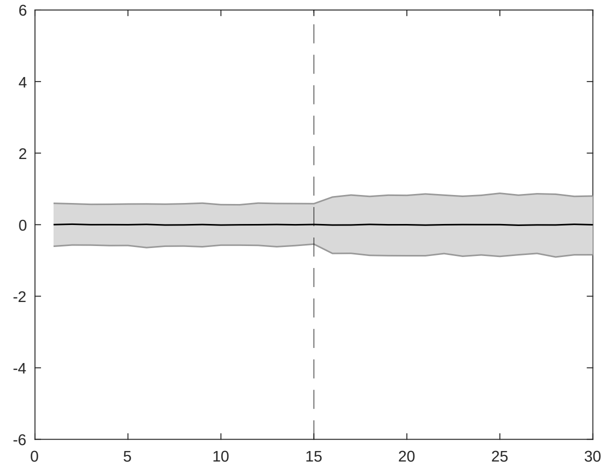
(d) $T_0 = 4$, $F = 100$ ($J = 199$)

Note: 95% bands for the simulation design of Figure 4.6.

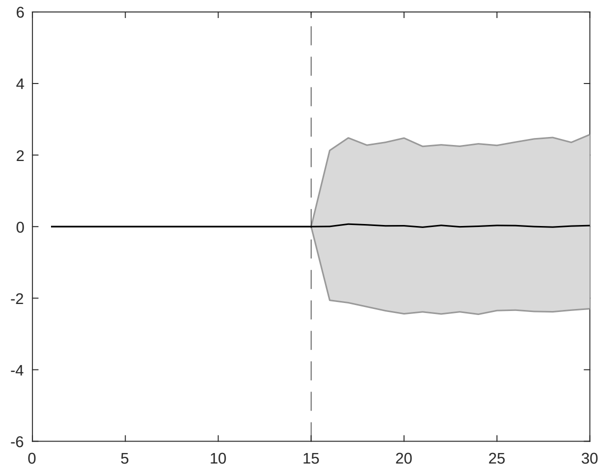
Figure A.7: Synthetic control error simulations for different weight restrictions.



(a) Standard SC



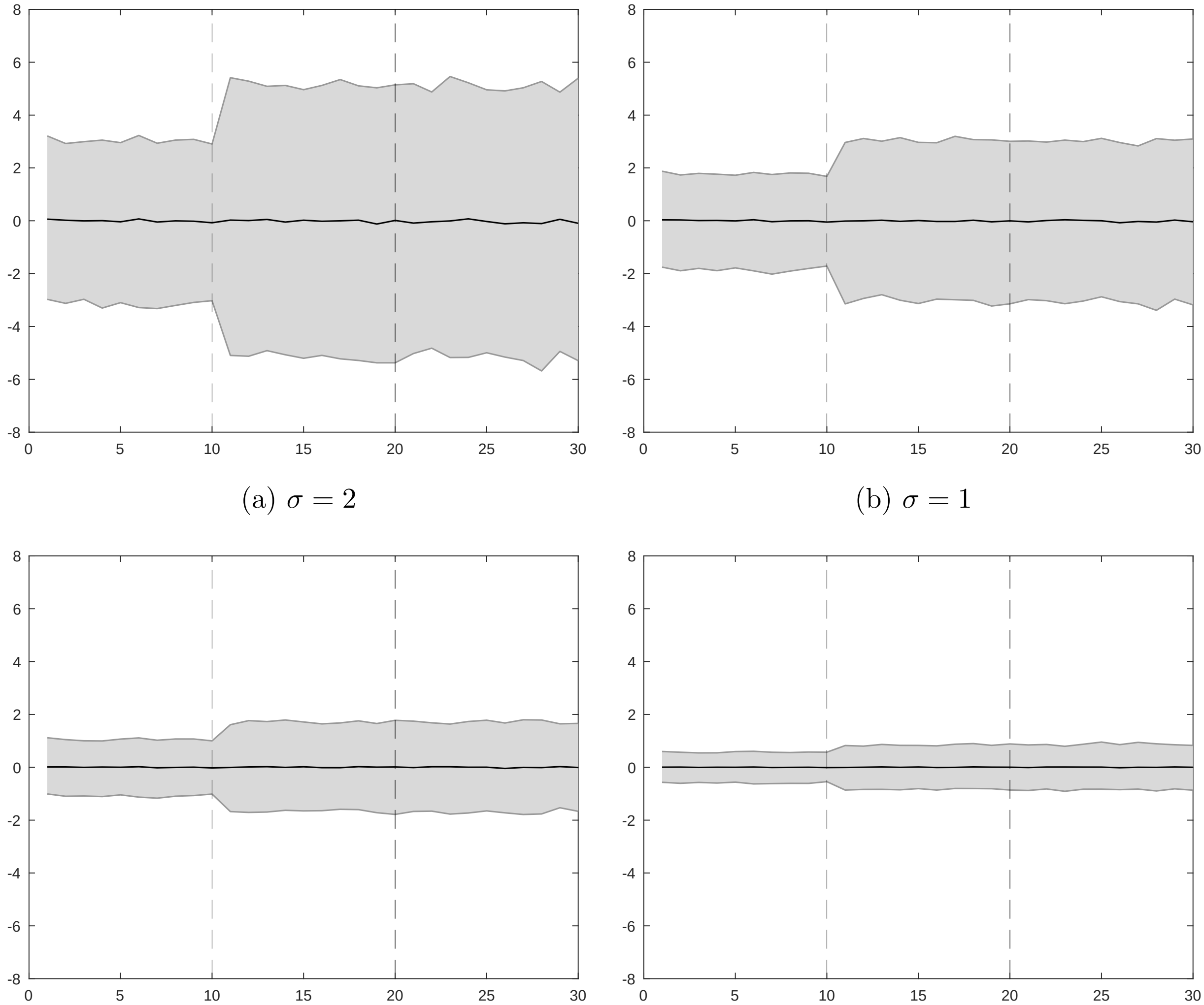
(b) SC with constant shift



(c) SC via regression

Note: 95% bands for the simulation design of Figure 4.7.

Figure A.8: Synthetic control error simulations with a validation period.



(a) $\sigma = 2$

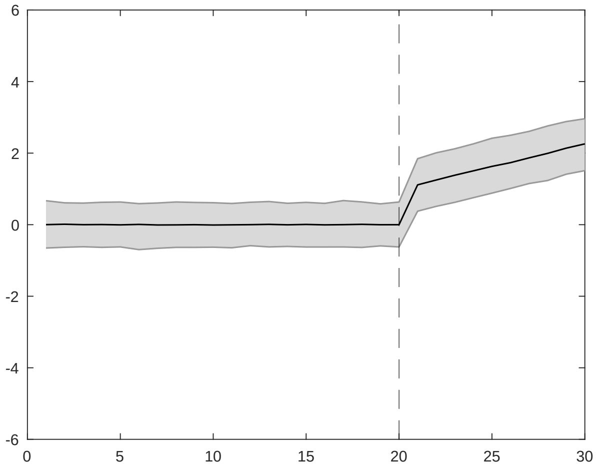
(b) $\sigma = 1$

(c) $\sigma = 0.5$

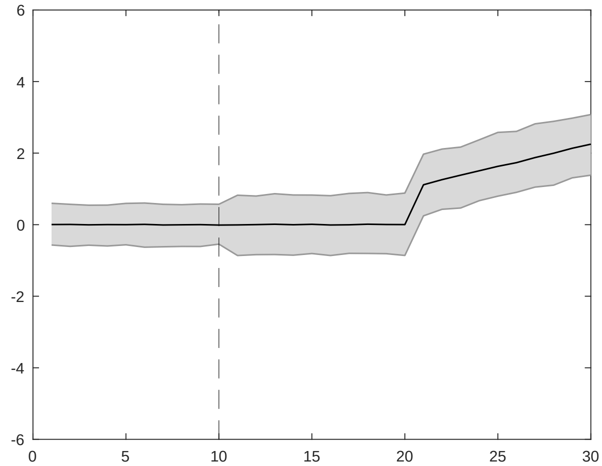
(d) $\sigma = 0.25$

Note: 95% bands for the simulation design of Figure 4.8.

Figure A.9: Synthetic control error simulations with treatment effect.



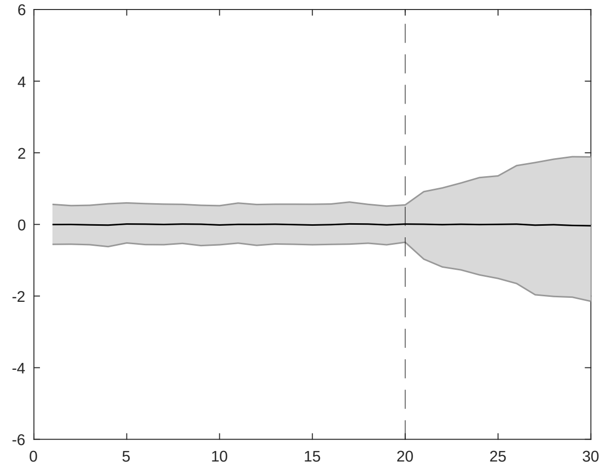
(a) Intervention at $T_0 = 20$



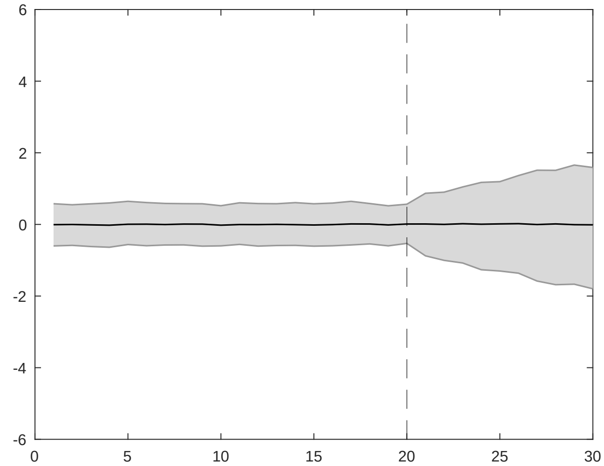
(b) Backdating to $T = 10$

Note: 95% bands for the simulation design of Figure 4.9.

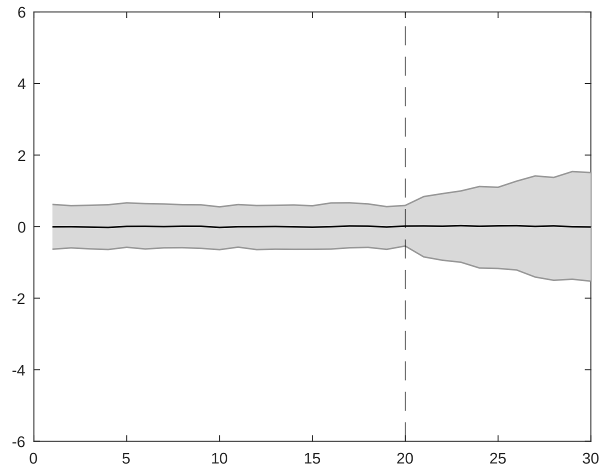
Figure A.10: Synthetic control error simulation with trimming.



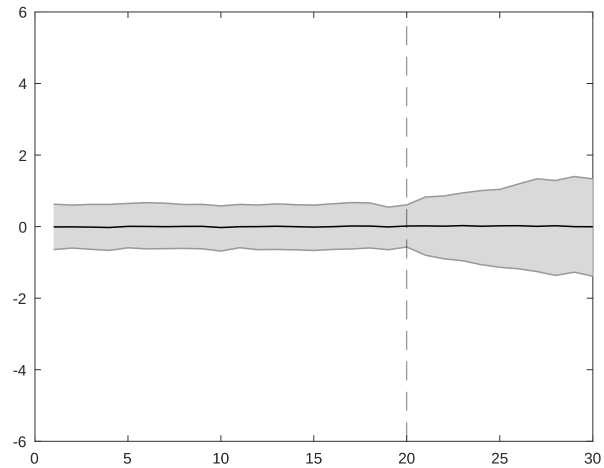
(a) Without trimming.



(b) 75% trimming.



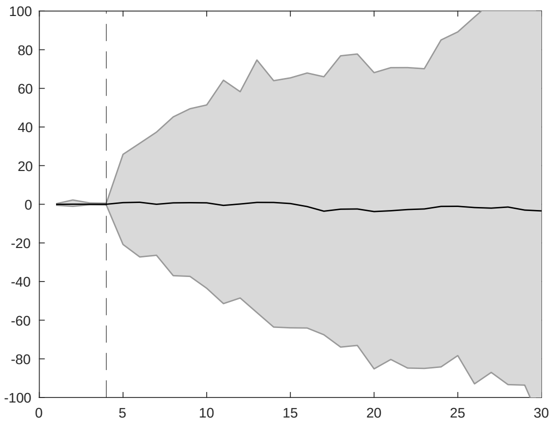
(c) 90% trimming.



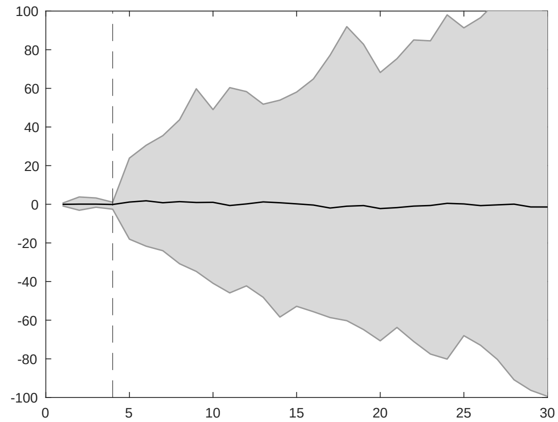
(d) 95% trimming.

Note: 95% bands for the simulation design of Figure 4.10

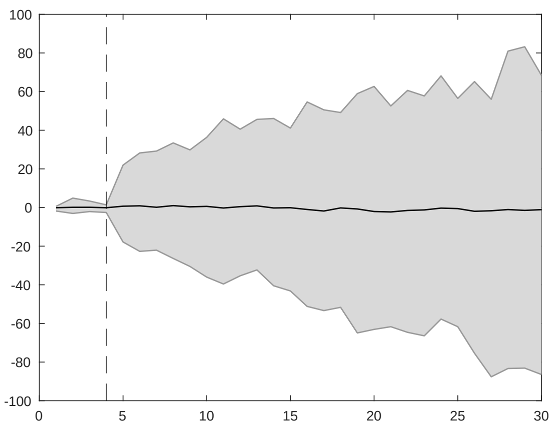
Figure A.11: Synthetic control error simulations with observed covariates.



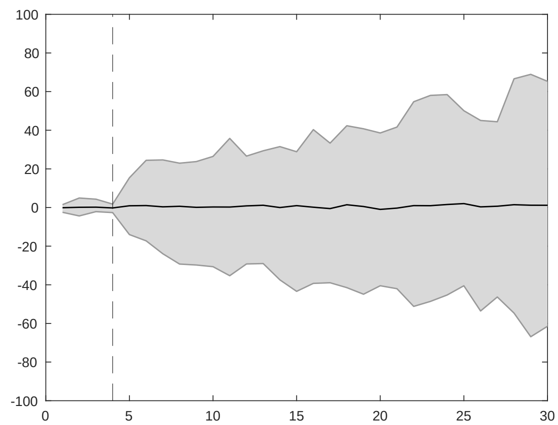
(a) $F = 5$



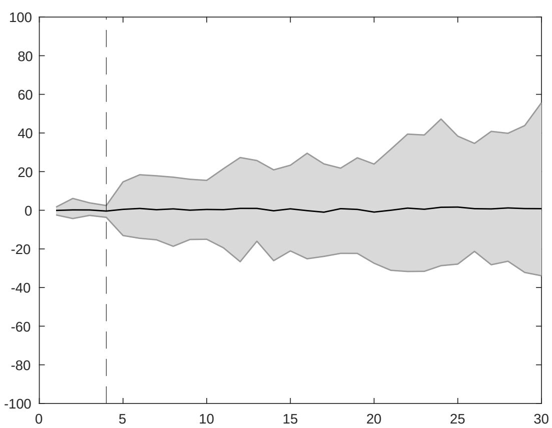
(b) $F = 4$



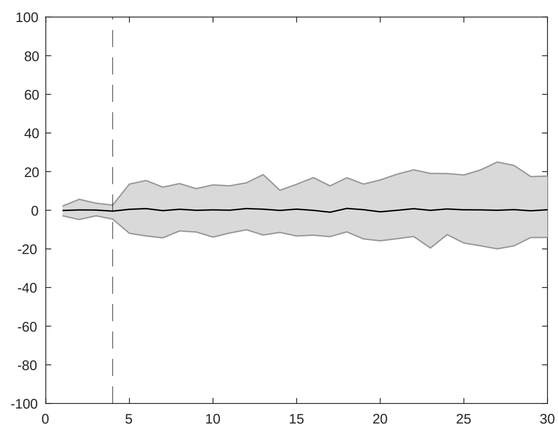
(c) $F = 3$



(d) $F = 2$



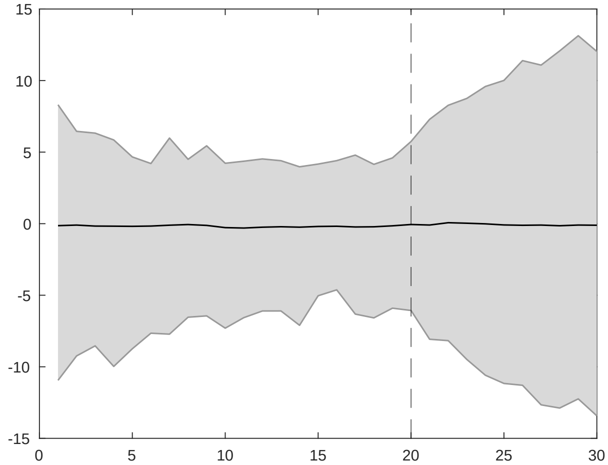
(e) $F = 1$



(f) $F = 0$

Note: 95% bands for the simulation design of Figure 4.11.

Figure A.12: Synthetic control error simulations with an auto-regressive process.



Note: 95% bands for the simulation design of Figure 4.12.

Bibliography

Abadie, Alberto, "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, 2021, *59* (2), 391–425.

_ **, Alexis Diamond, and Jens Hainmueller**, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.

_ **, _ , and _** , "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 2015, *59* (2), 495–510.

_ **and Guido W. Imbens**, "A Martingale Representation for Matching Estimators," *Journal of the American Statistical Association*, 2012, *107* (498), 833–843.

_ **and Jaume Vives-i-Bastida**, *Synthetic Controls in Action,* Econometric Society Monographs, Advances in Economics and Econometrics: Twelfth World Congress, Cambridge University Press, 2025.

_ **and Javier Gardeazabal**, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, March 2003, *93* (1), 113–132.

_ **and Jérémy L'Hour**, "A Penalized Synthetic Control Estimator for Disaggregated Data," *Journal of the American Statistical Association*, 2021. Forthcoming.

_ **and Jinglong Zhao**, "Synthetic Controls for Experimental Design," 2022.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton, "The value of connections in turbulent times: Evidence from the United States," *Journal of Financial Economics*, 2016, *121*, 368–391.

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales, "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Agarwal, Anish, Devavrat Shah, and Dennis Shen, "Synthetic Interventions," 2021. arXiv e-print, 2006.07691.

—, **Munther Dahleh, Devavrat Shah, and Dennis Shen**, "Causal Matrix Completion," 2021.

Aksu, Ege, Refik Erzan, and Murat Güray Kırdar, "The impact of mass migration of Syrians on the Turkish labor market," *Labour Economics*, 2022, p. 102183.

Amemiya, Takeshi, "Regression Analysis when the Dependent Variable Is Truncated Normal," *Econometrica*, 1973, *41* (6), 997–1016.

Amjad, Muhammad, Devavrat Shah, and Dennis Shen, "Robust Synthetic Control," *Journal of Machine Learning Research*, 2018, *19* (22), 1–51.

Anatolyev, Stanislav and Anna Mikusheva, "Factor models with many assets: Strong factors, weak factors, and the two-pass procedure," *Journal of Econometrics*, 2022, *229* (1), 103–126.

Angrist, Joshua D and Adriana D Kugler, "Protective or counter-productive? labour market institutions and the effect of immigration one unatives," *The Economic Journal*, 2003, *113* (488), F302–F331.

Arbour, David, Eli Ben-Michael, Avi Feller, Alex Franks, and Steven Raphael, "Using Multitask Gaussian Processes to estimate the effect of a targeted effort to remove firearms," 2021.

Arkhangelsky, Dmitry and David Hirshberg, "Large-Sample Properties of the Synthetic Control Method under Selection on Unobservables," 2023.

— **and Vasily Korovkin**, "On Policy Evaluation with Aggregate Time-Series Shocks," 2023.

—, **Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager**, "Synthetic Difference-in-Differences," *American Economic Review*, December 2021, *111* (12), 4088–4118.

Athey, Susan and Glenn Ellison, "Position Auctions with Consumer Search*," *The Quarterly Journal of Economics*, 08 2011, *126* (3), 1213–1270.

—, **Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi**, "Matrix Completion Methods for Causal Panel Data Models," *Journal of the American Statistical Association*, 2021. Forthcoming.

Autor, David H, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American economic review*, 2013, *103* (6), 2121–2168.

Bai, Jushan, "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 2009, *77* (4), 1229–1279.

Barron, Andrew R., "Entropy and the Central Limit Theorem," *The Annals of Probability*, 1986, *14* (1), 336 – 342.

Bartel, Ann P, "Where Do the New U.S. Immigrants Live?," *Journal of Labor Economics*, 1989, *7* (4), 371–91.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato, "Some new asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics*, 2015, *186* (2), 345–366.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein, "The Augmented Synthetic Control Method," *Journal of the American Statistical Association*, 2021, *116* (536), 1789–1803.

Berry, Steven, "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, *25* (2), 242–262.

Bonhomme, Stéphane and Elena Manresa, "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 2015, *83* (3), 1147–1184.

Borjas, George J, "The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market," *The quarterly journal of economics*, 2003, *118* (4), 1335–1374.

_, "Immigration economics," in "Immigration Economics," Harvard University Press, 2014.

_, "The wage impact of the Marielitos: A reappraisal," *ILR Review*, 2017, *70* (5), 1077–1110.

Borusyak, Kirill and Peter Hull, "Non-Random Exposure to Exogenous Shocks: Theory and Applications," NBER Working Papers 27845, National Bureau of Economic Research, Inc September 2020.

_, _, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

_, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," 2023.

Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott, "Inferring causal impact using Bayesian structural time-series models," *The Annals of Applied Statistics*, 2015, *9* (1), 247 – 274.

Card, David, "The impact of the Mariel boatlift on the Miami labor market," *ILR Review*, 1990, *43* (2), 245–257.

_, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

— **and Alan B. Krueger**, “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply,” *The American Economic Review*, 2000, *90* (5), 1397–1420.

Cengiz, Doruk and Hasan Tekgüç, “Is it merely a labor supply shock? Impacts of Syrian migrants on local economies in Turkey,” *ILR Review*, 2022, *75* (3), 741–768.

Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu, “A t-test for synthetic controls,” 2022.

— **, Kaspar Wüthrich, and Yinchu Zhu**, “An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls,” *Journal of the American Statistical Association*, 2021. Forthcoming.

Chetverikov, Denis, Zhipeng Liao, and Victor Chernozhukov, “On cross-validated Lasso in high dimensions,” 2016.

Choi, Hana and Carl F. Mela, “Monetizing Online Marketplaces,” *Marketing Science*, 2019, *38* (6), 948–972.

Compiani, Giovanni, Gregory Lewis, Sida Peng, and Peichun Wang, “Online Search and Optimal Product Rankings: An Empirical Framework,” *Marketing Science*, 2024, *43* (3), 615–636.

Cunningham, Scott and Manisha Shah, “Decriminalizing Indoor Prostitution: Implications for Sexual Violence and Public Health,” *The Review of Economic Studies*, 2018, *85* (3), 1683–1715.

Danieli, Oren, Daniel Nevo, Itai Walk, Bar Weinstein, and Dan Zeltzer, “Negative Control Falsification Tests for Instrumental Variable Designs,” *Working Paper*, 2024.

de Chaisemartin, Clément and Ziteng Lei, “More Robust Estimators for Instrumental-Variable Panel Designs, With An Application to the Effect of Imports from China on US Employment.,” 2023. Available at SSRN:https://ssrn.com/abstract=3802200.

— **, Xavier D’Haultfoeuille, Félix Pasquier, Doulo Sow, and Gonzalo Vazquez-Bare**, “Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period,” 2024.

Deaner, Ben, “Proxy Controls and Panel Data,” 2021.

Donohue, John J., Abhay Aneja, and Kyle D. Weber, “Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment Using Panel Data and a State-Level Synthetic Control Analysis,” *Journal of Empirical Legal Studies*, 2019, *16* (2), 198–247.

Doudchenko, Nikolay and Guido W Imbens, “Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis,” Working Paper 22791, National Bureau of Economic Research October 2016.

Esteller-Moré, Alejandro and Leonzio Rizzo, “The Economic Costs of a Secessionist Conflict: The Case of Catalonia,” *Defence and Peace Economics*, 2022, *33* (6), 655–688.

Ferman, Bruno, “On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls,” *Journal of the American Statistical Association*, 2021. Forthcoming.

_ and Cristine Pinto, “Synthetic Controls with Imperfect Pre-Treatment Fit,” *Quantitative Economics*, 2021. Forthcoming.

Firpo, Sergio and Vitor Possebom, “Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets,” *Journal of Causal Inference*, September 2018, *6* (2), 1–26.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro, “Pre-event Trends in the Panel Event-Study Design,” *American Economic Review*, September 2019, *109* (9), 3307–38.

Friedberg, Rachel M, “The impact of mass migration on the Israeli labor market,” *The Quarterly Journal of Economics*, 2001, *116* (4), 1373–1408.

Ghose, Anindya, Panagiotis G. Ipeirotis, and Beibei Li, “Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue,” *Management Science*, 2014, *60* (7), 1632–1654.

Gobillon, Laurent and Thierry Magnac, “Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls,” *The Review of Economics and Statistics*, 07 2016, *98* (3), 535–551.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, “Bartik instruments: What, when, why, and how,” *American Economic Review*, 2020, *110* (8), 2586–2624.

Grossmann, M., M. Jordan, and J. McCrain, “The Correlates of State Policy and the Structure of State Panel Data.,” *State Politics and Policy Quarterly*, 2021.

Gulek, Ahmet, “Formal Effects of Informal Labor: Evidence from Syrian Refugees in Turkey,” 2025. Available at: https://github.com/ahmetgulek/ahmetgulek.github.io/blob/main/FEoIL.pdf.

Hainmueller, Jens, “Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies,” *Political Analysis*, 2012, *20* (1), 25–46.

Hall, P. and C.C. Heyde, *Martingale limit theory and its application* Probability and mathematical statistics, Academic Press, 1980.

Ham, Dae Woong and Luke Miratrix, “Benefits and costs of matching prior to a Difference in Differences analysis when parallel trends does not hold,” 2022.

He, Xuming and Qi-Man Shao, "A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs," *The Annals of Statistics*, 1996, *24* (6), 2608 – 2630.

— **and** — , "On Parameters of Increasing Dimensions," *Journal of Multivariate Analysis*, 2000, *73* (1), 120–135.

Hodgson, Charles and Gregory Lewis, "You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Consumer Search," Working Paper 31697, National Bureau of Economic Research September 2023.

Hsiao, Cheng, H. Steve Ching, and Shui Ki Wan, "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," *Journal of Applied Econometrics*, 2012, *27* (5), 705–740.

Hunt, Jennifer, "The impact of the 1962 repatriates from Algeria on the French labor market," *ILR Review*, 1992, *45* (3), 556–572.

Imbens, Guido and Paul Rosenbaum, "Randomization Inference with an Instrumental Variable," *Journal of the Royal Statistical Society, Series A*, 2005, *168* (1), 109–126.

Imbens, Guido W. and Davide Viviano, "Identification and Inference for Synthetic Controls with Confounding," 2023.

— **and Donald B. Rubin**, "Bayesian inference for causal effects in randomized experiments with noncompliance," *The Annals of Statistics*, 1997, *25* (1), 305 – 327.

Jaeger, David A, Joakim Ruist, and Jan Stuhler, "Shift-Share Instruments and the Impact of Immigration," Working Paper 24285, National Bureau of Economic Research February 2018.

Jr, Melvin Stephens and Dou-Yan Yang, "Compulsory education and the benefits of schooling," *American Economic Review*, 2014, *104* (6), 1777–1792.

Kasy, Maximilian and Lukas Lehner, "Employing the unemployed of Marienthal: Evaluation of a guaranteed job program," *AEA RCT Registry*, 2021.

Kim, Sungjin, Clarence Lee, and Sachin Gupta, "Bayesian Synthetic Control Methods," *Journal of Marketing Research*, 2020, *57* (5), 831–852.

King, Gary and Langche Zeng, "The Dangers of Extreme Counterfactuals," *Political Analysis*, 2006, *14* (2), 131–159.

Kleven, Henrik Jacobsen, Camille Landais, and Emmanuel Saez, "Taxation and International Migration of Superstars: Evidence from the European Football Market," *American Economic Review*, August 2013, *103* (5), 1892–1924.

Klößner, Stefan, Ashok Kaul, Gregor Pfeifer, and Manuel Schieler, "Comparative politics and the synthetic control method revisited: a note on Abadie et al. (2015)," *Swiss Journal of Economics and Statistics*, 2018, *154* (11).

LaLonde, Robert J., "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review*, 1986, *76* (4), 604–620.

Lebow, Jeremy, "The labor market effects of Venezuelan migration to Colombia: reconciling conflicting results," *IZA Journal of Development and Migration*, 2022, *13* (1).

Li, Kathleen T., "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," *Journal of the American Statistical Association*, 2020, *115* (532), 2068–2083.

_ **and Garrett P. Sonnier**, "Statistical Inference for the Factor Model Approach to Estimate Causal Effects in Quasi-Experimental Settings," *Journal of Marketing Research*, 2023, *60* (3), 449–472.

Llull, Joan, "Immigration, Wages, and Education: A Labour Market Equilibrium Structural Model," *The Review of Economic Studies*, 09 2017, *85* (3), 1852–1896.

los Santos, Babur De and Sergei Koulayev, "Optimizing Click-Through in Online Rankings with Endogenous Search Refinement," *Marketing Science*, 2017, *36* (4), 542–564.

Malo, Pekka, Juha Eskelinen, Xun Zhou, and Timo Kuosmanen, "Computing Synthetic Controls Using Bilevel Optimization," WorkingPaper, MPRA December 2020.

Miao, W., Z. Geng, and E. Tchetgen Tchetgen, "Identifying Causal Effects With Proxy Variables of an Unmeasured Confounder," *Biometrika*, 2018.

Moehring, Alex, "Personalized Rankings and User Engagement: An Empirical Evaluation of the Reddit News Feed," OSF Preprints 8yuwe, Center for Open Science January 2024.

Mogstad, Magne and Alexander Torgovitsky, "Instrumental Variables with Unobserved Heterogeneity in Treatment Effects," Working Paper 32927, National Bureau of Economic Research September 2024.

Narayanan, Sridhar and Kirthi Kalyanam, "Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach," *Marketing Science*, 2015, *34* (3), 388–407.

Newey, Whitney K. and Daniel McFadden, "Chapter 36 Large sample estimation and hypothesis testing," in "in," Vol. 4 of *Handbook of Econometrics*, Elsevier, 1994, pp. 2111–2245.

Pang, Xun, Licheng Liu, and Yiqing Xu, "A Bayesian Alternative to Synthetic Control for Comparative Case Studies," *Political Analysis*, 2022, *30* (2), 269–288.

Peri, Giovanni and Vasil Yasenov, "The labor market effects of a refugee wave synthetic control method meets the mariel boatlift," *Journal of Human Resources*, 2019, *54* (2), 267–309.

Pinkney, Sean, "An Improved and Extended Bayesian Synthetic Control," 2021.

Pouliot, Guillaume A. and Zhen Xie, "Information Criteria and Degrees of Freedom for the Synthetic Control Method," December 2022.

Quistorff, Brian, Matt Goldman, and Jason Thorpe, "Sparse Synthetic Controls: Unit-Level Counterfactuals from High-Dimensional Data," *Microsoft Journal of Applied Research*, 2020, (5), 155–170.

Ray, Kolyan and Aad van der Vaart, "Semiparametric Bayesian causal inference," *The Annals of Statistics*, oct 2020, *48* (5).

Reimers, Imke and Joel Waldfogel, "A Framework for Detection, Measurement, and Welfare Analysis of Platform Bias," Working Paper 31766, National Bureau of Economic Research October 2023.

Roth, Jonathan, "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, September 2022, *4* (3), 305–22.

Rubin, Donald B, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 1974, *66* (5), 688.

Rudelson, Mark and Roman Vershynin, "Sampling from Large Matrices: An Approach through Geometric Functional Analysis," *J. ACM*, jul 2007, *54* (4), 21–es.

Rutz, Oliver J., Randolph E. Bucklin, and Garrett P. Sonnier, "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising," *Journal of Marketing Research*, 2012, *49* (3), 306–319.

Scott, Steven L and Hal R Varian, "Predicting the present with Bayesian structural time series," *International Journal of Mathematical Modelling and Numerical Optimisation*, 2014, *5* (1-2), 4–23.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199. Themed Issue: Treatment Effect 1.

Turkish Red Crescent and WFP, "Refugees In Turkey: Livelihoods Survey Findings," 2019. https://reliefweb.int/report/turkey/refugees-turkey-livelihoods-survey-findings-2019-entr.

UNHCR, "UNHCR refugee statistics," https://www.unhcr.org/refugee-statistics/ 2021. Accessed: 2022-06-15.

Ursu, Raluca M., "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 2018, *37* (4), 530–552.

van der Vaart, A. W., *Asymptotic Statistics* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1998.

Vives-i-Bastida, Jaume, "STRETCHING THE NET: MULTIDIMENSIONAL REGULARIZATION," *Econometric Theory*, 2021, p. 1–30.

_ , "Predictor Selection for Synthetic Controls," 2022.

_ **and Alejandro Sabal**, "The Effects of Regulating Food Delivery Platform Design," 2024. Working project.

Wolfers, Justin, "Did unilateral divorce laws raise divorce rates? A reconciliation and new results," *American Economic Review*, 2006, *96* (5), 1802–1820.

Zhao, Peng and Bin Yu, "On Model Selection Consistency of Lasso," *Journal of Machine Learning Research*, 2006, *7* (90), 2541–2563.